21st Century PSYCHOLOGY

A Reference Handbook

Edited by

William F. Buskist
Stephen F. Davis

21st Century PSYCHOLOGY

21st Century PSYCHOLOGY
A Reference Handbook

1

Edited by

Stephen F. Davis

Texas Wesleyan University

William Buskist

Auburn University

A SAGE Reference Publication

SAGE Publications

Los Angeles • London • New Delhi • Singapore

For information:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
33 Pekin Street #02-01
Far East Square
Singapore 048763

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

21st century psychology : a reference handbook / editors, Stephen F. Davis, William Buskist.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4129-4968-2 (cloth : alk. paper)
1. Psychology—Handbooks, manuals, etc. I. Davis, Stephen F. II. Buskist, William. III. Title: Twenty first century psychology.
BF131.A15 2008
150—dc22

2007031738

This book is printed on acid-free paper.

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

Publisher: Rolf A. Janke
Acquisitions Editor: Jim Brace-Thompson
Developmental Editor: Sara Tauber
Reference Systems Manager: Leticia Gutierrez
Production Editor: Belinda Thresher, Appingo
Cover Designer: Michelle Kenny
Marketing Manager: Amberlyn Erzinger

Contents

Volume One

Volume Two

PREFACE

Although human and animal behaviors have been topics of interest to scientists and others since antiquity, historians typically date the inception of modern psychology to the mid-1800s. More specifically, they have selected 1879, the year that Wilhelm Wundt established his experimental psychology laboratory at the University of Leipzig, as the year that modern psychology originated. At that time, Wundt believed that the goals of psychology were (a) to study "immediate" conscious experience using experimental methodology and (b) to investigate higher mental processes using nonexperimental techniques.

The change that psychology has undergone in the nearly 130 years since its founding has been nothing short of phenomenal. For example, the early portions of the 20th century witnessed the development and popularization of the now classic "schools of psychology" such as structuralism, functionalism, Gestalt psychology, and behaviorism. World War II and the Korean War spurred the development of modern clinical psychology. During the mid-1900s, individual schools rose to prominence and tended to dominate psychological research and theorizing. These dominant schools often clashed with clinical psychology. For example, disagreements between behaviorists and clinicians, which have their roots in the 1940s and 1950s, still persist.

Toward the end of the 1960s, the nature of the field began to change, and the face of modern psychology was forever altered. First, Ulrich Neisser's 1967 book, *Cognitive Psychology*, ushered in the "cognitive revolution" and put behaviorism on the decline. Technological advances in computer technology, which allowed researchers to simulate human thought and memory processes and to create images of neurological processes, played an inestimable role in modern psychology's metamorphosis. Likewise, advances in social concern and action increased psychologists' awareness of psychology's diversity and its ability to make significant contributions in these areas. To be sure, the face of contemporary psychology was changing drastically. In fact, in 1992 former American Psychological Association (APA) president George A. Miller believed that psychology had become "an intellectual zoo" (p. 40). Clearly, that situation has not changed, as psychology is evolving in the 21st century.

Nowhere are psychology's expansion and change seen more clearly than in the evolution of the APA. Founded in 1892 by G. Stanley Hall at Clark University in Worcester, Massachusetts, the APA began with 31 charter members. Currently, there are over 60,000 APA members and 56 divisions with which these members and other interested psychologists can affiliate. The diversity of the APA divisions clearly reflects the changing face of contemporary psychology. They include General Psychology (Division 1), the Study of Social Issues (Division 9), Clinical Psychology (Division 12), Pharmacology and Substance Abuse (Division 28), Mental Retardation and Developmental Disabilities (Division 33), Media Psychology (Division 46), International Psychology (Division 52), and Trauma Psychology (Division 56). Clearly, psychology in the 21st century continues to be a diverse and evolving field.

We have attempted to capture psychology's dynamic and evolving nature in the 14 sections and the 104 chapters in *21st Century Psychology.* Our cadre of authors is composed of both established professionals who have already left their marks on the field and aspiring young professionals whose imprints are certain to be recorded in subsequent years. We believe that our choice of traditional and cutting-edge topics reflects contemporary psychology's diverse nature. For example, our *traditional* topics include the following:

- Neurotransmission (Chapter 14)
- Traditional Neuroscience Research Methods (Chapter 15)
- Vision (Chapter 23)
- Perception (Chapter 27)
- Recent Trends in Classical Conditioning (Chapter 34)

Our *cutting-edge* topics include the following:

- Conducting Research on the History of Psychology (Chapter 5)
- Qualitative Research (Chapter 11)
- Imaging Techniques for the Localization of Brain Function (Chapter 16)
- Stimulus Equivalence (Chapter 39)
- Memory and Eyewitness Testimony (Chapter 47)

- Positive Psychology (Chapter 73)
- Human Performance in Extreme Environments (Chapter 74)
- Community Psychology (Chapter 94)

Whether the chapter deals with a traditional topic or a cutting-edge topic, you will find that the authors present the materials in a decidedly contemporary manner. We hope that you will enjoy reading the chapters in the *21st Century Psychology* as much as we have enjoyed assembling them for you.

No reference book is the product of the editors' and authors' efforts alone; the editorial and production teams at SAGE deserve special praise. We extend our special thanks to James Brace-Thompson (Senior Editor, SAGE Reference), Rolf Janke (Vice President and Publisher, SAGE Reference), Sara Tauber (Development Editor), Diana Axelsen (Senior Development Editor), Leticia Gutierrez (Systems Manager), Tracy Buyan (Reference Production Supervisor), and Belinda Thresher (Production Editor, Appingo).

Finally, we dedicate this book to all students of psychology; they truly are the promising future of our field.

Stephen F. Davis
William Buskist

REFERENCES AND FURTHER READINGS

Miller, G. A. (1992). The constitutive problem of psychology. In S. Koch & D. E. Leary (Eds.), *A century of psychology as science* (pp. 40–45). Washington, DC: American Psychological Association.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

About the Editors

Stephen F. Davis is an emeritus professor at Emporia State University. He served as the 2002–2003 Knapp Distinguished Professor of Arts and Sciences at the University of San Diego. Currently he is Visiting Distinguished Professor of Psychology at Texas Wesleyan University. In 2007, he was awarded the honorary Doctor of Humane Letters degree by Morningside College. Since 1966, he has published over 285 articles and 24 textbooks, and has presented over 900 professional papers; the vast majority of these publications and presentations include student coauthors. Among his recent books are *Handbook of Research Methods in Experimental Psychology, An Introduction to Statistics and Research Methods: Becoming a Psychological Detective* (with Randolph A. Smith), and *Handbook of the Teaching of Psychology* (with William Buskist). He has served as president of the Society for the Teaching of Psychology (Division 2 of the American Psychological Association), the Southern Society for Philosophy and Psychology, the Southwestern Psychological Association, and Psi Chi (the National Honor Society in Psychology). Davis received the American Psychological Foundation National Teaching Award and the Society for the Teaching of Psychology National Teaching Award. Additionally, he was selected as the first recipient of the Psi Chi Florence L. Denmark National Faculty Advisor Award. He is a fellow of the American Psychological Society and Divisions 1 (General), 2 (Society for the Teaching of Psychology), 3 (Experimental), and 6 (Behavioral Neuroscience and Comparative Psychology) of the American Psychological Association.

William Buskist is the Distinguished Professor in the Teaching of Psychology at Auburn University and a faculty fellow at Auburn's Biggio Center for the Enhancement of Teaching and Learning. In his 25 years at Auburn, he has taught over 32,000 undergraduates, mostly in large sections of introductory psychology. He serves as the section editor for the Generalist's Corner section of *Teaching of Psychology* and as a member of the National Institute on the Teaching of Psychology (NITOP) planning committee. Together with Steve Davis, he has edited two volumes on the teaching of psychology: *The Teaching of Psychology: Essays in Honor of Wilbert J. McKeachie and Charles L. Brewer* (Erlbaum, 2003) and *The Handbook of the Teaching of Psychology* (Blackwell, 2005); with Barry Perlman and Lee McCann, he has edited *Voices of Experience: Memorable Talks from the National Institute on the Teaching of Psychology* (American Psychological Society, 2005). He has also coedited several electronic books for the Society of the Teaching of Psychology (http://teachpsych.org/resources/e-books/e-books.php). He has published over 30 books and articles on the teaching of psychology. In 2005, he was a corecipient (with Leanne Lamke) of Auburn University's highest teaching honor, The Gerald and Emily Leischuck Presidential Award for Excellence in Teaching. In addition, he was the American Psychological Association's 2005 Harry Kirke Wolfe lecturer. He also is a recipient of the 2000 Robert S. Daniel Teaching Excellence Award from the Society of the Teaching of Psychology. He is a fellow of American Psychological Association Divisions 1 (General Psychology) and 2 (Society for the Teaching of Psychology), and he is currently serving as president of the latter. His proudest career achievement is having five of his graduate students honored with national teaching awards.

About the Contributors

Brenda Anderson is an associate professor in biological psychology at Stony Brook University in Stony Brook, New York. Dr. Anderson received her MS degree in psychology from Emporia State University, and her PhD in biological psychology from the University of Illinois at Urbana-Champaign. She is interested in the influence of the environment and lifestyle on neural connections. Her research is designed to test whether learning, exercise, or stress can alter synapse numbers, regional volume, and metabolic capacity. She is also interested in environmental control over neuronal vulnerability to metabolic challenges. She uses animal models to explore causality between variables shown to be related in humans. Dr. Anderson's research relies on anatomical methodologies that span the subcellular to regional levels of analysis.

Francisco Arcediano completed his undergraduate and graduate education at the University of Deusto (Spain), where he obtained a PhD in psychology and conducted graduate research in both psychology and computer science. He then went on to become a postdoctoral fellow and research scientist at the State University of New York at Binghamton, where he further specialized in the areas of associative learning and memory in both humans and nonhumans. He is currently completing his graduate education in computer science and software engineering at Auburn University, where he joined the Intelligent and Interactive Systems Laboratory, and is conducting research on artificial intelligence, human-computer interactions, cognitive science, and learning technologies. He is an active researcher in the areas of human cognition, animal learning and cognition, and computer science, and has published multiple papers investigating these issues. During his academic career he has taught undergraduate and graduate courses in artificial intelligence, intelligence, learning and conditioning, memory, and personal computer applications.

Jeremy Atkinson is a PhD candidate in biopsychology at the State University of New York at Albany. His research interests include game theory, physical attractiveness in humans, primate genital morphology, and the difference between functionality and attractiveness in the human body, as well as the evolution of specific cognitive modules such as face recognition and human "irrationality." He is currently investigating osseous markers of attractiveness in the human body, comparative genital morphology in primates, and the attractiveness of specific proportions in the human body.

David B. Baker received his PhD in counseling psychology from Texas A&M University in 1988. Prior to coming to the University of Akron, he was a member of the psychology department at the University of North Texas, where he was active in child clinical research, training, and practice. As a historian of psychology, he teaches the history of psychology at the undergraduate and graduate levels. A contributing author to three books and more than 30 scholarly articles and book chapters, he maintains an active program of research on the rise of professional psychology in 20th-century America. In addition to being a fellow of the American Psychological Association, he serves on two editorial boards and is an elected member of the Association's Council of Representatives. Baker became the director of the Archives of the History of American Psychology in 1999.

Lewis Barker was a professor of psychology and neuroscience, as well as director of the PhD program in neuroscience, while at Baylor University from 1972 to 2000. Currently a professor of psychology at Auburn University, he teaches behavioral neuroscience, learning, and large sections of introductory psychology. He has edited two books on appetite and has written textbooks on learning and general psychology.

Kenneth Barron is currently an associate professor of psychology at James Madison University (JMU). He received his PhD in social/personality psychology from the University of Wisconsin, Madison, in 1999. While in graduate school, he developed a particular love for teaching students about research methods and statistics, and his early efforts were recognized with a number of graduate student teaching awards. At JMU, he continues to teach coursework in research methods and statistics. He also coordinates a unique, yearlong residential Psychology Learning Community program for incoming freshmen. He has received JMU's Outstanding Junior Faculty Award and the Provost Award for Excellence in Advising. He also has been nominated for Distinguished Teacher and Distinguished Research Awards. His research focuses on motivation, teaching pedagogy, and research method issues. His coauthors—Allison R. Brown, Theresa E. Egan, Christopher R. Gesualdi, and Kimberly A. Marchuk—were undergraduate students at JMU. All four were members of the Psychology Learning Community program, and all four received undergraduate psychology awards for Outstanding Achievement in Statistics and Research to recognize their early commitment and excellence in their methodology coursework in psychology. His student collaborators each made significant contributions to Chapter 7, Validity.

W. Robert Batsell earned bachelor's degrees in biology and psychology from Southern Methodist University, and his PhD in experimental psychology, under the direction of Dr. Wayne Ludvigson, from Texas Christian University. Currently he is the Kurt D. Kaufman Associate Professor and chair of psychology at Kalamazoo College in Kalamazoo, Michigan. He is a biopsychologist whose teaching interests include general psychology, experimental psychology, psychology of learning, and biopsychology. His research focuses on the learning mechanisms that underlie taste and odor aversions in humans and nonhumans. His interest in this topic of study emerged from his undergraduate days, when he worked with Dr. Michael Best at SMU.

Ludy T. Benjamin, Jr., is Presidential Professor of Teaching Excellence and professor of psychology at Texas A&M University. An experimental psychologist by training, he began his academic career at Nebraska Wesleyan University, served 2 years as director of education for the American Psychological Association, and then joined the faculty at Texas A&M, where he has been for 27 years. Benjamin has received numerous teaching awards from Texas A&M University. His national teaching awards include the Distinguished Teaching in Psychology Award from the American Psychological Foundation and the Distinguished Career Contributions to Education and Training Award from the American Psychological Association. Benjamin's scholarly work includes 22 books and more than 150 articles and book chapters, most on the history of psychology, focusing on the early American psychology laboratories and organizations, the origins of applied psychology, and the popularization of psychology, including a concern with the evolution of psychology's public image and the public's understanding of the science and practice of psychology. His recent books include *A Brief History of Modern Psychology* (Blackwell, 2007), *A History of Psychology in Letters* (Blackwell, 2006), and *A History of Psychology: Original Sources and Contemporary Research* (Wiley-Blackwell, 2008).

Diane S. Berry is a senior research scientist at Children's Medical Center Dallas. She earned her AB in psychology from Colby College, and her PhD in social/developmental psychology from Brandeis University. Until recently, she served as a tenured professor of psychology at Southern Methodist University, where her work focused on social perception and interaction, personality and social behavior, social and perceptual development, close relationships, and evolutionary psychology. In 2005, she was recruited by Children's Medical Center Dallas to develop a research program focused on psychological influences on pediatric injury risk and strategies for child injury prevention. This work applies basic research in social, developmental, personality, cognitive, and perceptual psychology to the problem of pediatric injury and injury control. Dr. Berry's recent work has been supported by grants from the American Trauma Society, the National Institutes of Health, and the Children's Trust.

Kathryn Norcross Black learned to give intelligence tests as a graduate student at the University of Iowa. She was heavily involved in testing the gifted for either early admission to school or grade advancement while on the faculty in psychological sciences at Purdue University. She also worked as a psychological consultant for the Gifted Education Resource Institute at Purdue and on several occasions taught in their summer program for gifted students. Shortly after her academic retirement, she began working with the gifted program for the Barrington public schools. She also has consulted for the gifted programs in other private and public schools in suburban Chicago as well as for individual families of the gifted.

Roger K. Blashfield is a professor in the Department of Psychology at Auburn University. He earned his BS at Ohio State and his PhD at Indiana University. Before coming to Auburn, he was on the faculty at Pennsylvania State University and the University of Florida (psychiatry). His area of research interest is the classification of psychopathology. He has many publications in the area of classification and personality disorders, and was a member of the *DSM-IV* personality disorders workgroup.

Christopher M. Bloom earned his doctorate in behavioral neuroscience from Saint Louis University in 2001. While at Saint Louis University, he studied in the Animal Sleep Laboratory of Dr. A. Michael Anch, investigating the role of myelin degeneration in a number of behavioral systems, including sleep. The laboratory continues to investigate basic sleep science via animal models. Dr. Bloom is currently an assistant professor of psychology at the University of Southern Indiana and director of the Southern Indiana Affective Neuroscience Laboratory (SIAN). The SIAN lab researches the role of negative emotions such as fear and disgust in a number of neuropsychological domains.

Kent Bodily is a teaching fellow and doctoral candidate in the experimental psychology program at Auburn

University. His research falls into the areas of animal cognition and comparative psychology. His interests include concept learning, problem solving, conditioning, and spatial navigation. He uses the matching-to-sample (MTS) and same/different (S/D) tasks to determine the mechanisms underlying generalized matching and abstract-concept learning in pigeons. His research suggests that abstract-concept learning is driven by multiple-exemplar training in both MTS and S/D tasks. He also uses desktop-computer immersive digital environments to test the applicability of nonhuman research to human spatial navigation and problem solving. His research has demonstrated commonalities between pigeon and human navigation in analogous spatial navigation tasks.

Jennifer M. Bonds-Raacke is an assistant professor of psychology in the Department of Psychology and Counseling at the University of North Carolina at Pembroke. She obtained her PhD in experimental/cognitive psychology from Kansas State University in Manhattan, Kansas. Dr. Bonds-Raacke has taught classes on the topic of memory, and her research interests include autobiographical memories for media experiences and decision-making strategies of couples. In addition, Dr. Bonds-Raacke is a member of the Association for Psychological Science, the Midwestern Psychological Association, and Psi Chi. She is dedicated to the involvement of undergraduates in research endeavors.

Jay C. Brown was born in Wisconsin, receiving his BS and MS from the University of Wisconsin at Oshkosh. It was there that he first developed his love of the history of psychology, though he had developed a general love of history much earlier. He went on to the State University of New York at Stony Brook for his PhD. After doing postdoctoral training at Carnegie Mellon University, he was an assistant professor at Missouri State University. Currently he is an assistant professor at Texas Wesleyan University, where he regularly teaches the History of Psychology course.

Danny R. Burgess is an advanced doctoral student in the Department of Psychology at Auburn University. He earned his BS at the University of Southern Mississippi. His doctoral dissertation focuses on the clinical utility of the five-factor model versus Axis II of *DSM-IV-TR* versus Axis V of *DSM-IV-TR* when characterizing personality disorders. He plans to continue researching the classification issue of whether a categorical or dimensional model is the most precise and useful system in understanding psychopathology. He currently is completing his internship at the University of Wisconsin Health Sciences Center in the Departments of Psychiatry and Rehabilitation Medicine in Madison, Wisconsin.

Susan R. Burns is an associate professor in the Department of Psychology at Morningside College (Sioux City, IA). Although her teaching load is somewhat diverse in the area of psychology, her teaching emphasis is on the human development courses (i.e., child and adolescent psychology and developmental psychology). Beyond her course load, Susan actively engages students in the research process. She currently has several students involved in both group and individual research projects investigating a variety of topics (e.g., bullying in high school, empathy and gender roles in adolescents enrolled in an intake treatment facility, cross-sectional analysis of short-term memory functioning, academic dishonesty, and gender role beliefs in relationships). Her students consistently present at local, small and large regional, and national conferences. Dr. Burns is currently serving as a consulting editor for *Psi Chi Journal of Undergraduate Research*, and recently has assumed the position of managing editor-elect for the *Journal of Psychological Inquiry.* Dr. Burns received her BS and MS in experimental psychology from Emporia State University (Emporia, Kansas) and her PhD in personality/social psychology with an emphasis in child development from Kansas State University (Manhattan, Kansas).

Robert B. Cialdini received undergraduate, graduate, and postgraduate education in psychology at the University of Wisconsin, the University of North Carolina, and Columbia University, respectively. He is currently Regents' Professor of Psychology and Marketing at Arizona State University, where he has also been named Distinguished Graduate Research Professor. He has been elected president of the Society of Personality and Social Psychology and has been the recipient of the Distinguished Scientific Achievement Award of the Society for Consumer Psychology, the Donald T. Campbell Award for Distinguished Contributions to Social Psychology, and the Peitho Award for Distinguished Contributions to the Science of Social Influence. His interests in persuasion and social influence have manifested recently in an emphasis on consumer psychology, which he makes a large part of his graduate and undergraduate courses in interpersonal influence. His focus on the influence process is also evident in his projects, currently underway, to investigate the factors that incline people to behave according to the norms of the society, especially in the arena of environmental protection.

Andrew L. Cohen is a doctoral candidate in clinical psychology at Auburn University. Mr. Cohen graduated from UCLA and began his graduate training in 2002. His primary research interests include the developmental study of ADHD, with an emphasis in childhood ADHD, and other childhood externalizing disorders. He has earned departmental recognition for his excellence in teaching and has published in the literatures pertaining to the teaching of psychology and ADHD. Upon obtaining his doctorate, Mr. Cohen wishes to enter the professoriate as an instructor, a researcher, and a clinician.

William G. Collier earned his BS in psychology from Oklahoma Christian University. He earned his first master's degree in experimental psychology at the University of Central Oklahoma. He proceeded to earn another master's degree and a doctorate in experimental psychology, with an emphasis in cognition, at Texas Christian University. Dr. Collier is currently an assistant professor of cognitive psychology at the University of North Carolina at Pembroke. While in graduate school, Dr. Collier developed an interest in research questions that psychologists have grappled with throughout the history of modern scientific psychology, such as the nature versus nurture question. The nature versus nurture issue is also relevant to Dr. Collier's primary research interest in the psychology of music. Since music exists in multiple species (e.g., birds, whales, and humans) and in all known human cultures, it is likely that there is a significant biological component to music. However, the varieties of music seen across human cultures suggest a significant environmental influence as well. Future research should investigate how nature and nurture combine and interact to influence our experiences with music.

Thomas S. Critchfield is a professor of psychology at Illinois State University. His scholarly interests focus on the study of fundamental processes of operant learning in people, and the expression of these processes in complex everyday settings. He is a past president of the Association for Behavior Analysis International, a former associate editor of *Journal of the Experimental Analysis of Behavior*, and a fellow of the American Psychological Association (Division 25). In 18 years of university teaching, he received eight departmental awards for teaching and student mentoring and has published more than two dozen journal articles and book chapters with student coauthors. During that span, students working under his supervision have been honored for their scholarship by professional associations more than a dozen times.

Lisa Curtin is a professor of psychology at Appalachian State University and associate director for research at the ASU Institute for Health and Human Services. She routinely teaches abnormal psychology, psychotherapy courses, and seminars in addictive behaviors. Her research program actively involves undergraduate and graduate students, and focuses on the understanding and treatment of substance use disorders as well as other addictive disorders, primarily eating disorders. She received the Outstanding Young Faculty Member award in the College of Arts and Sciences, and she was named the Outstanding College of Arts and Sciences Advisor. She routinely partners with community agencies to conduct research and evaluate community health programs, including serving on the advisory council for the North Carolina Office on Disability and Health. She is actively involved in community tobacco prevention efforts and in the planning of a community integrative medicine center.

Alan M. Daniel is an experimental psychologist with a focus on animal learning. Daniel is presently working in the lab of Mauricio R. Papini at Texas Christian University. The overarching goal of his research is to advance the study of how behavior and ecology interact to produce adaptive behavioral responses, which is key to understanding the driving forces in evolution. Currently, his research is centered upon exploring the adaptive significance and origins of frustration. By examining the effects of drugs, artificial selection, and various other manipulations on successive negative contrast, Daniel seeks genetic components and developmental pathways involved in the emergence of frustration.

John M. Davis is a professor of psychology at Texas State University, San Marcos. He received his PhD in experimental-social psychology from the University of Oklahoma in 1974. He has worked as a psychologist and lived in Germany, China, England, and the United States. He has conducted research and published in the areas of international psychology, interpersonal and intergroup relations, refugee stress and adaptation, and health psychology. Recent publications include a chapter (1999) on health psychology in international perspective, an article (2000) on international psychology in the *Encyclopedia of Psychology* (APA/Oxford University Press), a chapter (2002, republished in 2004) on countering international terrorism from the perspective of international psychology, and several articles on psychology throughout the world in the *International Journal of Psychology.*

James C. Denniston received his BA in psychology from New York University in 1992, his MA in psychology from Bucknell University in 1994, and his PhD from the State University of New York at Binghamton in 1999. He is currently an associate professor and assistant chair of the Department of Psychology at Appalachian State University. His current research focuses on elementary information processing in animals, including mechanisms involved in the extinction of conditioned fear and the role of context in modulating behavior.

Elizabeth B. Denny, a North Carolina native, earned a master's degree and a doctorate in psychology with a specialization in clinical psychology at the University of North Carolina at Greensboro. She completed her doctoral internship at the Veteran's Administration Medical Center in Salem, Virginia. She has a long-standing interest in cognitive aspects of clinical disorders, with a particular interest in memory and depression. She is currently a professor in the Department of Psychology and Counseling at the University of North Carolina at Pembroke, where she formerly served as the department chair and as the director of the service agency counseling program.

Wendy Donlin is an assistant professor in the Department of Psychology at University of North Carolina, Wilmington. She received her MS and PhD in experimental psychology from Auburn University. Prior to joining the faculty at UNCW, she completed a two-year NIH postdoctoral fellowship in the Behavioral Pharmacology Research Unit at the Johns Hopkins University School of Medicine. Her interests lie in human and animal models of illicit drug use, and the development of behavioral interventions for drug abuse.

Richard L. Doty, PhD, is the director of the University of Pennsylvania's Smell and Taste Center. He is an author or coauthor of over 350 professional publications, and a consultant to over 50 scientific journals. Among his numerous awards are the James A. Shannon Award from the National Institutes of Health (1996); the 2000 Outstanding Scientists of the 20th Century Award from the International Biographical Centre in Cambridge, England (1999); the Olfactory Research Fund's Scientific Sense of Smell Award (2000); the William Osler Patient-Oriented Research Award from the University of Pennsylvania (2003); the Society of Cosmetic Chemists' Service Award (2004); and the Association for Chemoreception Science's Max Mozell Award for Outstanding Achievement in the Chemical Senses (2005). He was elected to fellow of the American Association for the Advancement of Science in 2005.

Lennis Echterling is professor and director of counseling psychology at James Madison University. He has more than 30 years of experience in promoting resilience, particularly during crises and disasters. He has provided disaster intervention services across the country, including in Mississippi and Texas after Hurricanes Katrina and Rita. Following the 9/11 attacks, he worked as a Red Cross volunteer with survivors at the Pentagon. More recently, he was a crisis counselor after the shootings at Virginia Tech University. His books include *Crisis Intervention: Promoting Resilience and Resolution in Troubled Times*, *Thriving! A Manual for Students in the Helping Professions*, *Beyond Brief Counseling*, and *Becoming a Community Counselor*. Dr. Echterling has received the College Award for Distinguished Service, James Madison University's Distinguished Faculty Award, Virginia Counselors Association's Humanitarian and Caring Person Award, and the Counseling Vision and Innovation Award from the Association for Counselor Education and Supervision.

Nicole Else-Quest completed her PhD in developmental psychology at the University of Wisconsin, Madison, in 2006. She is assistant professor of psychology at Villanova University, where she teaches undergraduate and graduate courses in life-span development and the psychology of gender. Her previous research includes studies of sexuality development, gender differences in temperament, and the role of emotions in mathematics learning. Her current research focuses on gender differences in emotion and motivation involved in mathematics learning.

Martha Escobar completed her undergraduate education at the University of Deusto (Spain). She then went on to obtain MA and PhD degrees in cognitive and behavioral sciences from the State University of New York at Binghamton. Since then, she has been at Auburn University, where she is an associate professor of psychology. Her research interests are learning and memory in human and nonhuman animals, with a special interest in determining the underlying principles and applications of Pavlovian (classical) conditioning. She is currently investigating issues such as memory interference, representation of absent stimuli, temporal control of behavior, and causal inference. Her teaching background includes courses in introductory psychology, the psychology of learning, animal behavior, learning and conditioning, and specialized seminars on associative learning. She has published multiple papers and book chapters in the areas of learning and memory, and directed several graduate and undergraduate theses in those areas.

Jennifer Featherston is the director of a nonprofit that serves individuals with disabilities in Northwest Arkansas. She received her master's degree in rehabilitation counseling psychology from the University of Texas Southwestern Medical Center at Dallas and is a doctoral candidate in rehabilitation education and research at the University of Arkansas. She is certified as a Rehabilitation Counselor, Vocational Evaluation Specialist, and an Assisted Living Administrator. She is a member of Chi Sigma Iota Counseling Academic and Professional Honor Society International and the National Rehabilitation Association. She is interested in qualitative inquiry because it allows researchers to better understand the lived experiences of individuals in more diverse populations.

Larry Featherston is a research associate at the National Office of Research on Measurement and Evaluation Systems and an adjunct instructor at the University of Arkansas. He received his master's degree in rehabilitation counseling psychology from the University of Texas Southwestern Medical Center at Dallas and is a doctoral candidate in rehabilitation education and research at the University of Arkansas. His current research interest focuses on wage differences and wage discrimination experienced by individuals with disabilities. He is a certified Rehabilitation Counselor and Vocational Evaluation Specialist. He is currently a member of the National Rehabilitation Association, National Rehabilitation Counseling Association, National Council on Rehabilitation Education, Society for the Teaching of Psychology, American Statistical Association, and American Educational Research Association. He is interested in disability studies because of his lifelong involvement in working with individuals with disabilities within the community.

Daniel Fienup is a doctoral candidate in school psychology at Illinois State University. At the time this chapter was published, he was completing a predoctoral internship at the May Institute in Boston. Fienup earned his BA in psychology from Washington University in St. Louis, where his interest in behavioral psychology was sparked. Following this experience he earned a master's degree from the behavior analysis and therapy program at Southern Illinois University in Carbondale, where his interest in studying stimulus equivalence began. This interest spans both basic and applied questions. Currently his research focuses on applying technology based on stimulus equivalence to the acquisition of real-life skills and on comparing the efficiency of stimulus equivalence methodology to that of traditional teaching methods.

Michael Firment received his MA in 1987 and his PhD in 1990 from the University of Cincinnati. His major field of study was cognitive psychology, with minors in perception and human factors. He became a member of the psychology faculty at Kennesaw State University in 1989. His dissertation topic was memory for conceptual categories as instantiated by proverb meanings. He is currently interested in techniques of maximizing retention and transfer of academic knowledge. Dr. Firment has taught perception since 1988.

Krista K. Fritson earned her doctoral degree at Forrest Institute of Professional Psychology (PsyD) in Springfield, Missouri, in 1997 with a major in clinical psychology. En route to her PsyD, Dr. Fritson earned her MS in clinical psychology from Fort Hays State University and her BS in general psychology from the University of Nebraska at Kearney. Dr. Fritson joined the University of Nebraska at Kearney as an assistant professor after serving as a part-time lecturer for 2 years. At UNK, Dr. Fritson teaches the clinical track courses, abnormal psychology, and general psychology while conducting research. Additionally, Dr. Fritson continues to work as a clinical psychologist. She serves as the supervising practitioner for a residential treatment facility for young boys, provides clinical supervisions for therapists in the community, and maintains a small private practice. Though Dr. Fritson provides clinical consultation and services to all age groups and many diagnoses, her specialization has been working with children, adolescents, and families. With over 21 years of experience working in the mental health field, Dr. Fritson has had the opportunity to provide services to many youth with autistic spectrum disorders.

Matthew W. Gallagher is a graduate student in the clinical psychology program at the University of Kansas, Lawrence. He received his MA in 2006 from the University of Kansas under the mentorship of C. R. Snyder. His research is currently focused on examining how hope, optimism, and other character strengths lead to the development and maintenance of flourishing mental health, as well as how the findings of positive psychology can be applied to the practice of clinical psychology. He is also currently serving as the managing editor for *The Handbook of Positive Psychology* (2nd ed., Oxford).

Gordon G. Gallup, Jr., is an evolutionary psychologist and a professor in the Department of Psychology at the State University of New York at Albany. He is a former editor of the *Journal of Comparative Psychology.* His current research focuses on the impact of evolution on human behavior, and includes work on genital morphology, brain evolution, paternal assurance tactics, rape prevention strategies, semen chemistry, and the relationship between body configuration and behavior. He is also known for work on mirror self-recognition, localizing self-awareness in the brain, research on schizophrenia as a self-processing deficit, and predator-prey relations. More detailed information about his work can be found at http://www.evolutionary-psych.com.

E. Scott Geller, Alumni Distinguished Professor, is director of the Center for Applied Behavior Systems at Virginia Polytechnic Institute and State University (Virginia Tech), where he has been a faculty member since 1969. The author of 31 books, 44 book chapters, and more than 300 research articles addressing the development and application of behavior-change interventions, he has helped to improve the quality of work life across the United States and in several other countries. Since 1990, his monthly articles on the psychology of safety in *Industrial Safety & Hygiene News* have reached audiences at more than 75,000 companies and, more recently, health-care facilities. Scott Geller's caring, dedication, talent, and energy have helped him earn a teaching award in 1982 from the American Psychological Association, as well as every university teaching award offered at Virginia Tech. In 2001, Virginia Tech awarded Dr. Geller the University Alumni Award for Excellence in Research. In 2002, the University honored him with the Alumni Outreach Award for his exemplary real-world applications of behavioral science, and in 2003 Dr. Geller received the University Alumni Award for Graduate Student Advising. In 2005, E. Scott Geller was awarded the statewide Virginia Outstanding Faculty Award by the State Council of Higher Education.

Peter J. Giordano is a professor and chair of psychology at Belmont University in Nashville, Tennessee, where he began his career in 1989. He received his BA, MA, and PhD (clinical psychology) from the University of North Carolina at Chapel Hill. Personality psychology is one of the courses that first stimulated his interest in pursuing graduate study. He has taught the personality psychology course regularly during his career, typically adopting a writing-intensive format. He has served as National Past President of Psi Chi (the National Honor Society in Psychology) and as the methods and techniques editor for the journal *Teaching of Psychology.* He is a fellow of the Society for the Teaching of Psychology, Division Two of the American Psychological Association. In 1996, he received the Chaney Distinguished Professor Award, the highest teaching honor on his campus, and this past year was nominated for the CASE Professor of the Year Program. He has also received awards from Belmont's Department of Athletics and the Division of Student Affairs for his support of students in these programs.

C. James Goodwin is an emeritus professor at Wheeling Jesuit University, where he taught for 30 years before taking an early retirement. He is currently a visiting professor at Western Carolina University and living in the mountains at the western edge of North Carolina. He earned a bachelor's degree from the College of the Holy Cross and a master's and PhD in experimental psychology from Florida State University, specializing in memory and cognition. He is a fellow of the American Psychological Association in Divisions 2 (teaching) and 26 (history), and he is a past president of Division 26. His research interests on the empirical side are in the area of cognitive mapping and way finding, but his prime interest is in the early history of experimental psychology in the United States. He is the author of two undergraduate textbooks, one in research methods (*Research in Psychology: Methods and Design*) and one in the history of psychology (*A History of Modern Psychology*).

Dana Gresky is an assistant professor at Texas Wesleyan University in Fort Worth, Texas. She is a recent graduate from Texas Christian University, where she completed her degree in experimental social psychology and met her husband. Rather than follow her own interests, she was motivated to learn more about the world of sport psychology because of her husband's enthusiasm about athletic performance.

Cathy A. Grover received her PhD in experimental psychology with an emphasis in behavioral neuroscience from Texas A&M University, College Station, Texas, in 1992. She currently is an associate professor and the director of the graduate experimental psychology program at Emporia State University, where she studies the behavioral effects of drugs of abuse on learning, memory, and social behaviors using rats as an animal model. Cathy enjoys mentoring graduate and undergraduate students in the Davis Laboratory. In 1990, Grover was awarded the Teaching Excellence Early Career Award, presented by Division 2 of the American Psychological Association. She teaches drugs, brain and behavior; physiological psychology; sensation and perception; theories of motivation; research methods and statistics; and foundations of psychology.

Sharon E. Guttman is an assistant professor in the Psychology Department at Middle Tennessee State University. Her research explores human visual perception, specifically focusing on the binding problem—the question of how the human brain represents connections between different aspects of a perceptual experience. Her current research also includes projects on cross-modal perception, examining the integration of visual information with information gleaned from the other senses, and interactions between cognitive processing and balance.

Diane F. Halpern is professor of psychology and director of the Berger Institute for Work, Family, and Children at Claremont McKenna College. She has won many awards for her teaching and research, including the 2002 Outstanding Professor Award from the Western Psychological Association, the 1999 American Psychological Foundation Award for Distinguished Teaching, and the Outstanding Alumna Award from the University of Cincinnati. Halpern was 2004 president of the American Psychological Association. In addition, Halpern has served as president of the Western Psychological Association, the Society for the Teaching of Psychology, and the Division of General Psychology of the American Psychological Association. She has published over 350 articles and many books, including *Thought and Knowledge: An Introduction to Critical Thinking* (4th ed.), *Sex Differences in Cognitive Abilities* (3rd ed.), and she is currently working on the 3rd edition of *Psychological Science* with Michael Gazzaniga and Todd Heatherton. She is also working with Fanny Cheung from Chinese University on a cross-cultural book, based on more than 60 interviews with women in powerful leadership positions with substantial family responsibilities, titled *Women at the Top: How Powerful Leaders Combine Work and Family.*

Jeffrey L. Helms is an associate professor in the Department of Psychology at Kennesaw State University. He earned his MA and PsyD in clinical psychology from Spalding University in Louisville, Kentucky. He completed his predoctoral internship at Jefferson County Internship Consortium and his postdoctoral training at Family Connections/Seven Counties Services. Prior academic appointments included teaching not only at the undergraduate level but also at the master's, specialist, and doctoral levels. He has previously practiced in California and Kentucky and currently holds licensure as a psychologist in Georgia. His publications, research, and practice interests are predominantly in the areas of adolescence and forensic psychology. He is an associate editor of the *Journal of Forensic Psychology Practice.* He has practiced in multiple settings including within the community mental health arena, where his client base was predominantly minority juveniles (and their families) who had become entangled in the criminal justice system. Dr. Helms is a member of the American Psychological Association and maintains a small, private, forensic and clinical consulting practice.

Lisa Hensley is an assistant professor of psychology at Texas Wesleyan University. Her interest in death and bereavement is an outgrowth of both her personal experiences and her teaching and research experiences in the psychology of aging.

Michelle Hernández is a graduate student in experimental psychology at Auburn University, adjunct faculty at Alabama State University, and member of the teaching fellows program at Auburn University. Her approach toward teaching and research is based on the belief that excellence does not depend solely on the number of things we know, but on how we approach the things we do not know. She has taught cognitive psychology and developmental psychology, and assisted in a variety of courses and laboratories. In class she focuses on presenting tasks that reinforce curiosity and sharing among students as part of the scientific discourse. This approach forms the base of the laboratory training that undergraduate students receive under her mentorship and assisting in her research experience. Her research interests are currently focused on navigation of two-dimensional computerized environments by pigeons. She is currently a coadministrator of psycCOMM, an online community for the exchange of ideas in psychology, including research, teaching, and the application of psychological findings to the community.

E. Benjamin H. Heuston is a doctoral candidate at Brigham Young University in Provo, Utah. His interests include education and learning theory, learner modeling, cognition, artificial intelligence, psychometrics and evaluation, delivery systems, and evolutionary and comparative functional morphology. He is the president and chief operating officer of the nonprofit Waterford Research Institute, which seeks to bring equity and excellence in education through the use of technology.

Rick Hollings received his PhD in psychology from North Carolina State University. He is a licensed practicing psychologist with Wake Forest University Hospital and works at the Fayetteville Family Life Center, one of the faith-based counseling centers under the auspices of the hospital. He has an interest in the integration of psychology and theology and is currently a student at Campbell University Divinity School, pursuing a master of divinity degree.

Cooper B. Holmes is professor of psychology at Emporia State University, Roe R. Cross Distinguished Professor, and a licensed psychologist. He was a research assistant for one year at the Menninger Foundation following his BA from Washburn University, obtained an MA in clinical psychology from Bowling Green State University, and completed the PhD in counseling from the University of Toledo. His clinical experiences include three years as a psychologist at Toledo State Hospital, then, along with teaching, consulting at a mental health center, consulting with a psychiatrist in private practice, and having a private office for diagnostic evaluations. While a professor, he received training in neuropsychology at the VA Medical Center in Kansas City, Missouri. Professional interests include diagnostic fads, suicide, ethics, misdiagnosis of brain disorders as psychiatric conditions, coping with real-life crises, and older literary accounts of "modern" psychological concepts.

Jeffrey Holmes has been an assistant professor of psychology at Ithaca College for four years since finishing his doctorate in counseling psychology. He teaches large sections of introductory psychology and a senior seminar on controversial issues in psychological theory and research. He also teaches a unique laboratory course that provides introductory students with a vital foundation in psychological research, and advanced students with the opportunity to conduct laboratory-training sessions. He has received departmental recognition for superior teaching, and has twice been nominated for the Ithaca College Teaching Excellence Award. His research is primarily in the areas of teaching of psychology and social psychology. The latter area provided the background for his group processes chapter in the current volume.

Hui-Chin Hsu is an associate professor in the Department of Child and Family Development at the University of Georgia. Her research interest is on social interactions between mothers and their infants. One of her research topics is whether and how mothers' emotional experience in parenting, such as separation anxiety, contributes to their sensitivity and responsiveness to infants' social signals during social interaction. She is also interested in investigating the outcomes of mother-infant interaction that seek to understand the linkage of the quality of social interaction between mother and infant to later child compliance and social understanding in toddlerhood and preschool age.

Tim Huelsman is an associate professor of psychology at Appalachian State University and director of the industrial-organizational psychology and human resource management program at Appalachian, a multidisciplinary program offering the master's degree. He has served on the Education and Training Committee of the Society for Industrial and Organizational Psychology, and is the current chair of the North Carolina Industrial-Organizational Psychologists. His specialty is organizational psychology, with special interest in personality and mood in the organization, organizational assessment, and organizational culture and climate. His consulting work is broad, but generally focuses on organizational assessment, organizational development, and program evaluation. He most enjoys opening students' eyes to the field of I-O psychology and participating in their development as professionals who merge research and practice. He resides in Boone, North Carolina, with his wife, Jeanie, and their son Calvin.

Matthew T. Huss is an associate professor at Creighton University in Omaha, Nebraska. He is a graduate of the clinical psychology training program (forensic emphasis) and the law and psychology program at the University of Nebraska-Lincoln. He is an author of over 40 different publications, including a forthcoming textbook on forensic psychology, *Forensic Psychology: Research, Practice, and Applications.* His research interests generally revolve around risk assessment in specific populations (e.g., domestic violence and sex offenders), pedagogy and psychology, and the law in general.

Mary Inman is an associate professor at Hope College. She has taught social psychology, including social cognition, for 15 years as a college professor. She has published several research articles on social cognition topics such as perceiving discrimination and forming impressions from word-of-mouth communication. For example, she studies how targets of discrimination and third-party observers decide whether an event reflects discrimination. Her research examines why high- and low-status groups perceive discrimination differently, how the body responds when a person perceives discrimination, and which interventions can bridge perceptual differences between different groups. She has been an invited reviewer for several leading social psychology journals, including *Journal of Personality and Social Psychology, Journal of Experimental Social Psychology*, *Social Cognition*, and *Basic and Applied Social Psychology.* Her upcoming book, *The Research Companion for Social Psychology,* teaches undergraduate college students how to conduct social psychology research. She resides in Holland, Michigan, with her husband and son.

Jessica G. Irons is an assistant professor of psychology at James Madison University. She earned her PhD in experimental psychology from Auburn University and an MS from Augusta State University. Her scholarship of teaching interests focuses on empirically supported teaching methods, specifically methods for teaching critical thinking.

John Juve currently serves as a statistical analyst for JurySync, a litigation consulting firm in Olathe, Kansas. At JurySync, he conducts psychological and legal research investigating juror attitudes, beliefs, and knowledge, and their influence on decision making in the courtroom. He also developed and maintains a proprietary item bank, or virtual library, consisting of hundreds of items and questions used to support his jury selection research. Before coming to Kansas, he spent four years at the University of Missouri-Columbia completing his doctoral degree in educational psychology. He specialized in measurement, assessment, and statistics pertaining to the fields of psychology and education. While at the University of Missouri-Columbia, he taught several undergraduate and graduate courses and served as a research assistant to the associate dean for undergraduate programs and teacher development. He was selected to participate in a summer internship with the Psychological Testing Corporation in San Antonio, Texas, where he conducted psychometric research pertaining to the Wechsler Intelligence Scale for Children.

James Kalat is professor of psychology at North Carolina State University, where he has taught introduction to psychology and biological psychology since 1977. He received an AB degree summa cum laude from Duke University in 1968 and a PhD, under the supervision of Paul Rozin, from the University of Pennsylvania in 1971. He is the author of *Biological Psychology* (9th ed., 2007) and *Introduction to Psychology* (8th ed., 2008), and coauthor with Michelle N. Shiota of *Emotion* (2007), all published by Wadsworth. In addition, he has written journal articles on taste-aversion learning, the teaching of psychology, and other topics.

Jeffrey S. Katz is an alumni associate professor in the Department of Psychology at Auburn University. His research focus is in the area of comparative cognition. Ongoing projects involve avian same/different concept learning, the mechanisms by which pigeons learn matching to sample, behavioral mechanisms of auditory and visual list memory in primates, and problem solving in virtual environments. He teaches undergraduate and graduate classes in advanced experimental psychology, animal learning and cognition, cognitive psychology, and sensation and perception. He also currently holds teaching and research grants from the National Science Foundation and National Institutes of Mental Health. He has been honored with the Young Investigator Award—APA's Division 3 Experimental Psychology (2001), Outstanding Professor—Auburn University Panhellenic Council (2003), Psi Chi Excellence in Undergraduate Teaching—Department of Psychology Auburn University (2002), and College of Liberal Arts Early Career Teaching Award—Auburn University (2004–2005).

Jared W. Keeley is an advanced doctoral student in the Department of Psychology at Auburn University. He earned a BA degree from Knox College in Illinois. His research has investigated if clinicians use a hierarchical model to classify mental disorders, as is implied in the *DSM*, as well as how clinicians conceptualize comorbidity. Keeley is also actively interested in scholarship on the teaching of psychology, and plans to pursue an academic career.

Kenneth D. Keith is professor of psychology at the University of San Diego, where he was department chair from 1999 to 2007. He teaches coursework and a laboratory in cross-cultural psychology and conducts research on cross-cultural quality of life. He has a special interest in Japan and Japanese culture. In addition to his cross-cultural work, he has written extensively on other aspects of quality of life, intellectual disability, and the teaching of psychology.

Marcel Satsky Kerr is associate professor of psychology at Texas Wesleyan University and serves as chair of the Department of Psychology. Kerr has taught psychology at the university level for over 10 years at four different schools in Texas—Texas Tech University, South Plains College, Texas Wesleyan University, and Tarleton State University. She received her bachelor's in psychology from Texas A&M University, a MA in experimental psychology from Texas Tech University, a PhD in experimental psychology and statistical methods from Texas Tech University, and a MEd in educational technology from the University of Texas at Brownsville. Kerr's interest in psychometrics emerged during her doctoral work, when survey development projects presented opportunities for skill development and needed income. Her current research focuses on online learning pedagogy, which allows her to continue honing her psychometric skills.

Beverly R. King is an associate professor in the Psychology and Counseling Department at the University of North Carolina at Pembroke. She holds a PhD in developmental psychology from Purdue University and teaches a wide variety of psychology courses, both face-to-face and online. Some of these courses are introductory psychology, child and adolescent psychology, cross-cultural child development, psychology of gender, and human development and personality. Her interest in repressed and recovered memories stems from her use of the topic in many of these courses to illustrate how psychological research can address important societal issues.

Mary Kite received her BA, MS, and PhD from Purdue University. She is now professor of psychological science at Ball State University. Recently, she served in administrative roles at Ball State, including Acting Graduate Dean and Associate Graduate Dean. Throughout her career, she has maintained an active research program in the area of stereotyping and prejudice, particularly as it relates to antigay prejudice and racism. She has published in a number of journals, including *American Psychologist, Psychological Bulletin, Journal of Personality and Social Psychology, Personality and Social Psychology Bulletin, Psychology of Women Quarterly,* and *Sex Roles.* Most recently, she has coauthored a textbook, *The Psychology of Prejudice and Discrimination,* with Bernard Whitley. She is a fellow of the Society of the Teaching of Psychology, the Society for the Psychology of Women, and the Society for the Psychological Study of Lesbian, Gay, and Bisexual Issues. In 2006, she served as president of the Society for the Teaching of Psychology. She is currently secretary-treasurer of the Midwestern Psychological Association and chairs the APA Task Force on Diversity Education Resources.

Stephen B. Klein has been professor and head of the Department of Psychology at Mississippi State University since 1990. Dr. Klein obtained his BS degree from Virginia Tech in 1968 and his PhD from Rutgers University in 1971. He taught for 12 years at Old Dominion University and was chair of the Department of Psychology at Fort Hays State University for 7 years prior to coming to Mississippi State University. Dr. Klein has written numerous articles on the biological basis of learning and memory and is the author of seven textbooks, including *Learning: Principles and Applications* (2002, 4th ed., McGraw Hill, adopted by several hundred universities and translated into Spanish) and *Biological Psychology* (2007, Worth, coauthored with B. Michael Thorne). He also coedited the two-volume text *Contemporary Learning Theories* (2001), both published by Lawrence Erlbaum.

Christopher Koch is a professor of psychology at George Fox University, where he has been department chair, director of scholarship, and director of assessment. He teaches Sensation & Perception, Neuroscience, Cognition, Research Methods, and Statistics. He earned his PhD in cognitive-experimental psychology from the University of Georgia and studies perceptual and attentional processes. Dr. Koch has also conducted cross-cultural research in perception and cognition as a Fulbright scholar in Russia. He is coauthor of CognitionXL, which is an online educational system for active learning about sensory, perceptual, and cognitive processes. Dr. Koch has been president of Psi Chi (the National Honor Society in Psychology) and a Councelor for the Psychology Division of the Council for Undergraduate Research.

Philip Kortum is currently a professor in the Department of Psychology at Rice University in Houston, Texas. Prior to joining Rice, he worked for almost a decade at SBC Laboratories (now AT&T Laboratories) doing human factors research and development in all areas of telecommunications. Dr. Kortum continues to do work in the research and development of user-centric systems in both the visual (Web design, equipment design, image compression) and auditory domains (telephony operations and interactive voice response systems). He received his PhD in biomedical engineering from the University of Texas at Austin.

John H. Krantz is professor of psychology at Hanover College. He received his undergraduate degree from St. Andrews Presbyterian College and his master's and doctorate from the University of Florida. His training is in human vision. The principal research methods he was taught and uses are psychophysical methods. Immediately after receiving his doctorate, he worked for Honeywell on cockpit display development. In this work he was able to learn and see how psychophysics contributes to applied research. Since coming to Hanover College in 1990, he has taught a wide range of courses including sensation and perception, biopsychology, cognition, and research methods. All of these courses are laboratory courses. In teaching these courses, particularly the laboratories, he has focused on conveying to his students the importance of clearly understanding one's research methods in how they frame both research questions and answers. His interest in the nature of research methods, such as psychophysics, has led to his work in the exploration of the use of the World Wide Web for psychological research.

David Kreiner is professor of psychology and associate dean of the Graduate School at the University of Central Missouri, where he began teaching in 1990 after completing a PhD in human experimental psychology at the University of Texas, Austin. He enjoys teaching research design & analysis, cognitive psychology, sensation & perception, advanced statistics, and a colloquium on the psychology of language. Research interests include language processing, memory, and the teaching of psychology. He often collaborates with undergraduate and graduate students on publications and presentations. He is a member of the American Psychological Association Divisions 2 (Teaching of Psychology) and 3 (Experimental Psychology), the Association for Psychological Science, the Midwestern Psychological Association, and the Council of Teachers of Undergraduate Psychology.

Jason Kring is an assistant professor of human factors and systems at Embry-Riddle Aeronautical University in Daytona Beach, Florida. He received his MS in experimental psychology from Emporia State University, and his PhD in applied experimental and human factors psychology from the University of Central Florida. He has worked as a researcher at the United States Army Research Institute in Orlando and interned at NASA's Johnson Space Center in Houston, developing human factors recommendations related to the International Space Station. He is currently President of the Society for Human Performance in Extreme Environments (HPEE); an interdisciplinary forum for scientists, operational personnel, and students with an interest and expertise in the area of human performance and behavior in complex, high-stress environments. He also serves as editor of the *Journal of Human Performance in Extreme Environments* and is codirector of the Team Simulation and Gaming Laboratory at Embry-Riddle. He has over 30 publications and 50 conference presentations, representing a wide range of research interests including performance in extreme and stressful settings; effects of cohesion, trust, and communication on team performance; crew composition and interactions for long-duration spaceflight; aerospace human factors; distributed team performance; and training applications of simulation and virtual reality.

Natalie Kerr Lawrence received her PhD in experimental social psychology from Virginia Commonwealth University in 2001. She became interested in the study of attitudes during graduate school and completed her dissertation on social identity processes in attitude change. She is now an assistant professor in the Department of Psychology at James Madison University, where she teaches general psychology, statistics, research methods, social psychology, and social influence. She is an active member of the Society for the Teaching of Psychology, and her current research interests are focused on the teaching of psychology

Steven R. Lawyer is a clinical psychologist and an assistant professor in the Department of Psychology at Idaho State University. He received his BS from Western Michigan University in 1995 and his PhD from Auburn University in 2002. His research interests include experimental psychopathology, sexual decision making, and sexual aggression. His clinical work primarily concerns the treatment of anxiety and mood-related disorders using empirically supported treatments such as cognitive-behavioral therapies. He teaches graduate and undergraduate courses concerning psychotherapy and psychological science.

Michael J. T. Leftwich earned his PhD in clinical psychology from Oklahoma State University in 1999, with a subspecialty emphasis in health psychology. He is currently at Forest Institute of Professional Psychology. His primary interests include substance-related disorders, personality theory and assessment, and cognitive-behavioral therapy.

Philip Lewis is a professor of psychology at Auburn University. The recipient of numerous departmental, college, and university teaching awards, Lewis teaches graduate and undergraduate courses on Personality Theory and Adolescent and Adult Development. Phil's interest in a life-span approach to personality development began when he was a doctoral student at Syracuse University working under the direction of the cognitive developmental psychologist David E. Hunt. Phil was the lead investigator in a longitudinal study of West Point cadets and is currently involved in a longitudinal study of transformational change in Harvard MBA students.

Carolyn A. Licht received her doctorate in clinical psychology with a child and family specialization from Fordham University. She recently became nationally certified as an Acupuncture Detoxification Specialist (ADS). Previously she had an accomplished 10-year professional ballet career dancing in New York, Iowa, Wisconsin, and Illinois; in 1992, she developed a creative dance and movement course for children with special needs. In 2006, she coauthored two published articles introducing self-report instruments designed to measure the Bowen Family Systems constructs of pursuer-distancer and differentiation of self. She is currently employed at the Family Care Center at Harlem Hospital in New York City.

Shane J. Lopez is associate professor of counseling psychology at the University of Kansas, Lawrence, where he teaches courses in positive psychology, psychological assessment, and educational leadership. He also is a Gallup senior scientist, a role through which he consults primarily with the Gallup Education Division and Gallup University. He serves on the editorial board of *The Journal of Positive Psychology* and on the advisory board for Ready, Set, Learn, the Discovery Channel's preschool educational television programming. Through his current research programs, Lopez is examining the effectiveness of hope training programs in the schools (under the auspices of the Making Hope Happen Program), refining a model of psychological courage, and exploring the link between soft life skills and hard outcomes in education, work, health, and family functioning. His books include *The Handbook of Positive Psychology* (Oxford) and *Positive Psychological Assessment: A Handbook of Models and Measures* (American Psychological Association Press), both with C. R. Snyder.

Denise Martz is a professor and coordinator for graduate programs in psychology at Appalachian State University.

She regularly teaches health psychology, behavioral medicine, women's health, and large sections of abnormal psychology. She most enjoys mentoring student research, and has chaired numerous master's and honors theses and assisted many students in publishing their thesis work. Her research team's studies on "fat talk," the social psychology of body image, have received widespread media attention. In 2000, she was the recipient of the Graduate Student Association Award for Outstanding Advising and Mentoring. She has engaged in the private practice of behavioral medicine and psychotherapy in her community for the past 12 years. Eager to bring health psychology to mainstream medicine and public health, she serves a leadership role in planning an integrative medicine center, a partnership between her university and local hospital system, in Boone, North Carolina.

George Mather is professor of experimental psychology in the Department of Psychology at the University of Sussex, Brighton, United Kingdom. He received his bachelor's degree in psychology from Sheffield University, United Kingdom, and his PhD in visual psychophysics from Reading University, United Kingdom. He was a postdoctoral fellow at Reading University, United Kingdom; York University, Canada; and University College London, United Kingdom, before joining the faculty at Sussex University. Mather has published numerous research papers on a range of topics, including human spatial vision, depth perception, and motion perception. He is the joint editor (with Frans Verstraten and Stuart Anstis) of *The Motion Aftereffect: A Modern Perspective* (MIT Press, 1998), and the author of *Foundations of Perception* (Psychology Press, 2006). He is currently engaged in research into human visual motion perception, and teaches undergraduate courses in perception and in the relation between art and psychology.

Janet R. Matthews is a tenured professor in the psychology department at Loyola University, New Orleans, where she has taught for well over 20 years. She is a licensed and board-certified clinical psychologist. Dr. Matthews received her PhD in clinical psychology from the University of Mississippi following her predoctoral clinical internship at the University of Oklahoma Health Sciences Center. Her postdoctoral fellowship in clinical and neuropsychological assessment was at the University of Nebraska Medical Center. She has written numerous articles about ethics as well as other areas of clinical practice. Her clinical psychology textbook was recently published. She served a term on her state psychology licensing board as well as various APA governance groups, including its Council of Representatives and Board of Directors. Her husband, Lee, is also a licensed and board-certified clinical psychologist. Outside psychology, she enjoys spending time with her cats and reading mystery novels.

Kurt D. Michael is an associate professor of psychology at Appalachian State University (ASU) and Associate Director of Clinical Services at the ASU Institute for Health and Human Services. He teaches abnormal psychology and history and systems at the undergraduate level, and child psychopathology and interventions for children and adolescents at the graduate level. His primary area of research is treatment outcome for child and adolescent disorders as well as substance abuse disorders. He is presently involved in a longitudinal project designed to investigate the efficacy of motivational interviewing in reducing problematic drinking patterns in first-year college students. Thus far, the results appear promising, and a subset of these data was published recently in the American Psychological Association journal, *Professional Psychology: Research and Practice.* In addition to Michael's teaching and research interests, he is a practicing licensed psychologist and oversees the outcome evaluation process for the largest private addictive treatment center in the Carolinas.

Harold L. Miller, Jr., is professor of psychology and Karl G. Maeser Professor of General Education at Brigham Young University. He is associate editor of the *Journal of the Experimental Analysis of Behavior* and has published research reports and reviews in that journal and others devoted to behavior analysis. His primary research interests are choice and decision making in humans and other species, with particular attention to qualitatively different reinforcers, delay discounting, and self-control.

Richard L. Miller received his PhD in 1975 from Northwestern University. He has taught at Georgetown University and the University of Cologne. He served for many years as the director of applied behavioral science research for HumRRO in Heidelberg, Germany, where he was involved in research on leadership, organizational behavior, and environmental and community psychology. Since 1990, he has held the position of professor and chair of the Psychology Department at the University of Nebraska at Kearney. He is a recipient of the University´s Outstanding Teaching and Instructional Creativity Award, a past president of RMPA, and a fellow of both APA and APS.

Adriana Molitor is an assistant professor in the Department of Psychology at the University of San Diego. She received her baccalaureate degree from the University of California at Riverside, with a double major in psychology and human development, and earned her doctorate in developmental psychology from Duke University. Before joining the faculty at USD, she completed a postdoctoral research position at the Yale University Child Study Center. Dr. Molitor's teaching interests include infancy, child and adolescent development, and developmental research methods. Her research interests center around the social and emotional development of infants and toddlers. The topics of her research include the dynamics of mother-infant interactions, the impact of these interactions on emotional and social behavior, the problems encountered by at-risk infants and toddlers, and the role of culture in structuring developmental goals and mother-child exchanges.

David Morgan is a full professor in the School of Professional Psychology at Spalding University in Louisville, Kentucky. Over the course of the past 16 years, he has taught nearly a dozen different courses in both the undergraduate psychology program and the graduate MA and PsyD programs. He maintains specific scholarly interests in behavior analysis and single-subject research designs, and has published numerous articles and books in these areas. A number of years ago, he developed an undergraduate course titled *Environmental Problems and Human Behavior.* Unlike most environmental psychology courses, *Environmental Problems and Human Behavior* focuses on human behavior as the independent variable and environmental quality as the dependent variable. Major themes in the course have been the poor fit between humans' Stone Age dispositions and the current high-tech world in which many of us live, as well as the basic application of fundamental behavior principles to the amelioration of environmental problems.

Robin K. Morgan is a professor of psychology at Indiana University Southeast in New Albany, Indiana. Over the course of the past 18 years, she has taught over 20 different courses at the undergraduate level and was the founding director of the campus Institute for Learning and Teaching Excellence. Among her teaching awards are the Master Teacher Recognition Award from Indiana University Southeast, the 2001 Herman Fredric Lieber Memorial Award—Distinguished Teaching Award for Indiana University, the Indiana University Southeast Distinguished Teaching Award, membership in F.A.C.E.T. (Indiana University Faculty Colloquium on Excellence in Teaching), and the Metroversity Outstanding Adult Teacher Award. Her scholarly work has focused on both the scholarship of teaching and how individuals cope with stressful life events, and she has published articles, chapters, and books in these areas. Her current research is focused on student stalking of faculty.

Daniel D. Moriarty is professor and chair of the Department of Psychology at the University of San Diego and director of research at the California Wolf Center. His current interests include management of wolves in captivity and the use of conditioned taste aversion to control predatory behavior. As a comparative psychologist, he has long recognized evolutionary theory as the unifying principle in the understanding and explanation of behavior. Extension of the evolutionary perspective to human behavior has been challenging but fruitful, as the development of evolutionary psychology has clearly shown.

Chad Mortensen received his undergraduate degree in psychology from the University of Iowa. He is now a doctoral student in social psychology at Arizona State University. His interests include social influence and a functional approach to social psychology.

Sherman M. Normandin is a graduate student in the clinical psychology doctoral training program at Idaho State University. He received his BA at the University of North Dakota in 2000. His research interests include using experimental procedures to examine contextual factors that influence sexual decision making in a laboratory environment. His clinical interests include using behavioral principles to manage childhood disorders as well as adult psychopathology, including substance abuse and anxiety disorders. He is also interested in the study of the history of psychology in general and the history of clinical psychology in particular.

Joseph J. Palladino is professor of psychology and chair of psychology at the University of Southern Indiana (Evansville, Indiana). He earned his PhD in general theoretical psychology from Fordham University. Since 1981, he has been on the faculty of the University of Southern Indiana, where he has served as chair of the Psychology Department, coordinator of university assessment, and coordinator of programs for the university's Center for Teaching and Learning Excellence. He is a fellow of the American Psychological Association (APA) and its Division on Teaching and has served the Division as president (1991–1992), editor of the Methods and Techniques section of *Teaching of Psychology,* and chairperson of the program committee. In 2000, he was elected Midwestern vice-president of Psi Chi (the National Honor Society in Psychology). In 1982, he founded the Mid-America Undergraduate Psychology Research Conference, which is now run by a consortium of faculty representing colleges and universities in Indiana, Illinois, and Kentucky. In 1990, he received the Division 2 Teaching Excellence Award. He received the University of Southern Indiana Alumni Association's Faculty Recognition Award in 2000. His autobiography, "From the streets of the Bronx to academia," was selected to appear in *The Teaching of Psychology in Autobiography: Perspectives from Exemplary Psychology Teachers,* which was published in 2005 by the Society for the Teaching of Psychology.

Mauricio R. Papini, professor of psychology at Texas Christian University, has dedicated his career to the comparative analysis of learning in a variety of species, focusing on situations involving surprising omissions or reductions in incentive magnitude. His current research centers on the neurobiological basis of incentive contrast, with emphasis on limbic structures and the opioid system. The long-term goal of this research is to provide a better understanding of the evolution of brain mechanisms responsible for incentive contrast phenomena. Mauricio has published *Comparative Psychology: Evolution and Development of Behavior* (2002), translated into Japanese and Spanish and soon to appear in a second edition; was editor of the *International Journal of Comparative Psychology* (2000–2006); and is president-elect of the International Society for Comparative Psychology.

Sharon Pearcey received her master's and PhD degrees from Georgia State University in behavioral neuroscience.

During the 2000–2001 academic year, Pearcey completed a postdoctoral teaching fellowship in the Psychology Department at Furman University. She is currently an assistant professor of psychology at Kennesaw State University. Her research interests are in the areas of food intake, meal patterns, and activity. She has published her research in several journals, including *Physiology & Behavior, Appetite,* and *The American Journal of Clinical Nutrition.* Pearcey is a member of the Council on Undergraduate Research and is committed to working with undergraduate students on research projects.

Kerri Pickel is a professor and the director of graduate studies in the Department of Psychological Science at Ball State University in Muncie, Indiana. Her research focuses on psychology and law, including eyewitness memory and credibility. She has published articles investigating the weapon focus effect, differences between accurate and inaccurate witnesses, the consequences of lying on witnesses' memory, and how jurors evaluate witnesses' testimony. At Ball State, she has taught undergraduate courses in research methods, cognition, and statistics, as well as courses in cognition and memory for the graduate program in Cognitive and Social Processes. She is a member of the editorial board of *Law and Human Behavior* (the journal of Division 41 of the American Psychological Association).

Wade Pickren earned his PhD in the history of psychology with a minor in the history of science under the direction of Don Dewsbury at the University of Florida. He is the associate chair of the Department of Psychology at Ryerson University in Toronto, where he also leads the Cultural Strengths Research Group. For 8 years, Pickren was APA historian and director of archives (1998–2006). After moving to Ryerson in 2006, he remains APA Historian. He has two edited books, *Evolving Perspectives on the History of Psychology* and *Psychology and the NIMH*. His most recent book, *Psychology and the Department of Veterans Affairs: A Historical Analysis of Training, Research, Practice, and Advocacy*, was published in 2007. He has served as a guest editor and contributor to several issues of *American Psychologist*: "The Contributions of Kenneth and Mamie Clark," "Psychology and the Nobel Prize," and "50 years after *Brown v Board of Education*." Pickren's scholarly interests include the use of narratives in understanding cultural strengths, the history of psychology and health care, the history of efforts to make psychology truly inclusive in both theory and practice, and the history of indigenous psychologies. He enjoys classroom teaching and aims to help students develop a disciplined curiosity about psychology and life.

Loreto R. Prieto, PhD, is director of the U.S. Latino/a studies program and professor of psychology at Iowa State University. He is a fellow of the American Psychological Association (APA) through both the Society for the Teaching of Psychology (STP) and the Society of Counseling Psychology. Prieto has over 100 publications and presentations, many focusing on diversity issues, including the upcoming text *Got Diversity? Best Practices For Incorporating Culture into the Curriculum,* coedited with Dr. Regan Gurung. Prieto was the chair of the STP Diversity Task Force (1998–2005) and also served as chair of the Working Group on Diversity for the APA Psychology Partnerships Project (P3). He is a sought-after speaker and consultant on diversity issues, especially as they pertain to the teaching of psychology.

John Raacke is currently an assistant professor of psychology at the University of North Carolina at Pembroke (UNCP). Dr. Raacke received both his master's and PhD in experimental/cognitive psychology from Kansas State University. He is currently the chair of the Institutional Review Board (IRB) at UNCP and is a member of the Association of Psychological Science as well as the Society of Judgment and Decision Making. His research interests include team decision making, the use of statistical evidence in juror and jury decision making, and the longitudinal development of expertise. Dr. Raacke has published over 10 articles and 4 book chapters, and has written and received over 10 grants. In addition, he has taught classes in a variety of areas in psychology including sensation and perception, problem solving and decision making, physiological psychology, research methods, and developmental psychology.

Erin Rasmussen is an associate professor in the Department of Psychology at Idaho State University, where she teaches undergraduate and graduate courses in learning and behavioral pharmacology. She received her MS and PhD in experimental psychology (with a minor in behavioral pharmacology and toxicology) from Auburn University. She taught at the College of Charleston as an assistant professor for three years before joining the faculty at Idaho State University. Her past research involved examination of how prenatal exposure to heavy metals affects the behavior of offspring and the role of environmental enrichment in attenuating those effects. Currently, she is examining the behavioral economics of food choices that lead to obesity in environmental and genetic rodent models of obesity, as well as underlying neurotransmitter systems (endocannabinoids and opioids) that are involved in food reinforcement.

Shilpa Pai Regan is an assistant professor of clinical psychology at the University of North Carolina at Pembroke. In this position, she teaches undergraduate and graduate psychology courses in introductory psychology, adult development, multicultural counseling, and diagnosing and assessing individuals. She maintains an active research program focusing on mental health, specifically the appropriate treatment and assessment of multicultural populations. Dr. Regan is a graduate of the University of North Carolina at Chapel Hill, earned her master's and doctorate in clinical psychology from Oklahoma State University, and is licensed by the North Carolina Psychology Board. She specializes in cognitive-behavioral therapy, motivational interviewing, and multicultural counseling. She is experienced in working with adults, couples, families, children, adolescents, and groups. Dr. Regan has significant knowledge and expertise in the evaluation and treatment of substance abuse (alcohol and illicit drugs), trauma reactions, depression, and anxiety disorders.

Jessica M. Richmond is currently a graduate student at the University of Akron in the counseling psychology doctoral program. She received her master's in clinical psychology from Radford University in 2006. Her research interests include end-of-life care and decision making, quality of life in individuals living with HIV/AIDS, and long-term effects of childhood sexual abuse. She has been involved in several research projects related to HIV disease, specifically focusing on decision making in HIV-positive pregnant women, and has authored/coauthored several articles and book chapters on issues dealing with the end of life.

Verena M. Roberts is a graduate student in the clinical psychology doctoral training program at Idaho State University. She was born in Berlin, Germany, but has lived in the United States since 2001. Her research interests include the assessment and treatment of sexual dysfunction in relationships, a focus she plans to apply to her professional clinical work. Clinically, she employs an eclectic mix of therapeutic principles to individualize her treatments and meet her clients' needs, and embraces an evidence-based approach to psychological care that includes cognitive-behavioral therapy and principles.

Alexandra Rutherford is an associate professor of psychology at York University in Toronto. She is a primary faculty member in, and past coordinator of, the history and theory of psychology graduate program. She has served for the past five years as chair of the Heritage Committee for the Society for the Psychology of Women of the American Psychological Association. She also serves as the official historian of the Society for the Psychological Study of Social Issues. Her teaching and research interests and publications span a number of areas, including the history of women and feminism in psychology, the history and social impact of behavioral technologies, and the history of clinical psychology. From 2002 to 2005 she was the book review editor of the *Journal of the History of the Behavioral Sciences*, and since 2002 she has edited a regular column in *The Feminist Psychologist* highlighting the roles of women in the history of psychology. In addition to her work as an academician, she is also a practicing clinical psychologist.

Joseph J. Ryan received his PhD from the University of Missouri-Columbia and is currently professor and chair in the Department of Psychology at the University of Central Missouri, Warrensburg, Missouri. He is a diplomate in clinical neuropsychology of the American Board of Professional Psychology/American Board of Clinical Neuropsychology and a fellow of the American Psychological Association and the National Academy of Neuropsychology. Ryan was the 2006 recipient of the Raymond D. Fowler Award from the American Psychological Association of Graduate Students (APAGS) and, in 2007, was named the William F. Byler Distinguished Faculty at the University of Central Missouri. The author of more than 200 publications, he has received the Outstanding Graduate Professor of Psychology Award from UCM students each of the past five years. Prior to assuming his position at UCM in 1999, he was chief of the Psychology Service at the Dwight D. Eisenhower Veterans Affairs Medical Center and adjunct professor in the Department of Psychiatry at the University of Kansas Medical Center. He has served on the editorial boards of numerous professional publications, including *Archives of Clinical Neuropsychology, Journal of Psychoeducational Assessment*, and *The Clinical Neuropsychologist.*

Kimberly Rynearson is an associate professor of psychology at Tarleton State University. She holds a BA in psychology (The University of Texas at Austin), a MA in experimental psychology (Texas Tech University), a MEd in adult education and distance learning (University of Phoenix), and a PhD in experimental psychology with an emphasis in cognition (Texas Tech University). Dr. Rynearson's research interests include characteristics of successful online learners, the role of verbally immediate behavior in the online classroom, adult learning theory and its application in a computer-mediated learning environment, the use of technology to enhance K–12 reading instruction, word recognition and reading comprehension among beginning and skilled readers, and characteristics and practices of expert readers.

Bryan K. Saville is an assistant professor in the Department of Psychology at James Madison University (JMU) in Harrisonburg, Virginia, where he has been since the fall of 2004. Prior to joining the faculty at JMU, he was an assistant professor in the Department of Psychology at Stephen F. Austin State University in Nacogdoches, Texas. He earned a BA in psychology from the University of Minnesota, a MS in applied psychology (behavior analysis) from St. Cloud State University, and a PhD in experimental psychology from Auburn University. In 2002, he received the McKeachie Early Career Award from the Society for the Teaching of Psychology (STP; Division 2 of APA). He coedited "E-xcellence in Teaching," a monthly e-column devoted to the teaching of psychology, and *Essays from E-xcellence in Teaching* (Volumes 3–6), both of which are published on STP's Web site. He has published over 20 journal articles and book chapters, and is author of the forthcoming book, *A Guide to Teaching Research Methods in Psychology* (Blackwell).

Lauren F. V. Scharff is a professor in the Department of Psychology at Stephen F. Austin State University (SFASU). She completed her PhD in human experimental psychology in December 1992 from the University of Texas at Austin, and started teaching at SFASU in January 1993. She is codirector for the SFASU Teaching Excellence Center, which she helped establish in 2006. Dr. Scharff regularly teaches introductory psychology, research methods, biopsychology, perception courses, and a teaching seminar. Her major research interests include text readability, Web site usability, visual search, and depth perception, although her students continually shift her research efforts to new directions. For the past nine years, she has actively collaborated with researchers at NASA-Ames to create a metric to predict text readability and to study information processing in the visual system.

Kelly Schmidtke is a graduate student at Auburn University. At Auburn, she works as a teaching assistant and a research assistant in the Animal Learning and Cognition Laboratory. She is interested in using animal models to understand how organisms acquire skills. Using pigeons as subjects, her current research focuses on how expectancies influence acquisition of the same-different concept.

Steven K. Shapiro received his PhD in clinical psychology from the University of Miami in 1990 upon completing an internship at Duke University Medical Center. Currently, Dr. Shapiro is director of clinical training and associate professor of psychology at Auburn University, where he has been since 1990. His research activities focus on the description and assessment of executive functioning, learning, and conduct problems in children and adults. Dr. Shapiro has taught a variety of undergraduate and graduate courses, and is particularly interested in child psychopathology, child behavioral and cognitive assessment, and clinical supervision.

Jason Sikorski is a clinical psychologist privileged to be an assistant professor of psychology at Central Connecticut State University. His research focuses on disruptive behavior problems in adolescence and, more specifically, in predicting risk for sexual offense recidivism in juveniles. However, Dr. Sikorski takes the most pride in being an effective and passionate teacher of psychology. He is a recipient of the McKeachie Early Career Award given yearly by Division 2: The Society for the Teaching of Psychology of the American Psychological Association. This prestigious honor recognizes exemplary classroom teaching, scholarship pertaining to the teaching of psychology, and the establishment of a professional identity as a teacher of psychology.

Leah Skovran attends the University of Nebraska, Lincoln, as a PhD/MLS student in social psychology and law. Her general research interests include the broad area of legal decision making and, more specifically, the role of emotion in capital murder cases. Her research focus also includes a forensic clinical perspective, particularly in the area of psychopathy and sexual offenders, and their relationship to sensation-seeking behaviors.

Randolph A. Smith completed his undergraduate degree at the University of Houston and PhD at Texas Tech University in experimental psychology (specialties in human learning/memory and statistics). Randy taught at Ouachita Baptist University in Arkansas for 26 years, chaired Kennesaw State University's Psychology Department for 4 years, and became chair of Lamar University's Psychology Department in 2007. His professional work centers on the scholarship of teaching. Randy serves as editor of the Society for the Teaching of Psychology's journal *Teaching of Psychology.* He is author of *Challenging Your Preconceptions: Thinking Critically About Psychology* (2002), coauthor (with Steve Davis) of *The Psychologist as Detective: An Introduction to Conducting Research in Psychology* (2007), and coauthor (with Steve Davis) of *An Introduction to Statistics and Research Methods: Becoming a Psychological Detective* (2005). He has worked with high school teachers grading AP exams since the test's inception and recently served as faculty advisor for TOPSS (Teachers of Psychology in Secondary Schools). He is a member of the American Psychological Association and the Association for Psychological Science. In 2006, Smith received the American Psychological Foundation's Charles L. Brewer Distinguished Teaching of Psychology Award and the University System of Georgia Regents' Scholarship of Teaching and Learning Award.

Todd A. Smitherman is currently completing a two-year postdoctoral fellowship in the Department of Psychiatry and Human Behavior at the University of Mississippi Medical Center. He completed his PhD in clinical psychology at Auburn University, where his research focused on the assessment and treatment of anxiety disorders, particularly panic disorder, generalized anxiety disorder, and specific phobias. He has published numerous journal articles and book chapters on these topics. His current interests also include behavioral medicine and the comorbidity of anxiety and depressive disorders with various medical conditions, such as recurrent headache. In this regard, he has worked to increase awareness of anxiety and mood disorders in medical patients and to develop screening instruments to improve assessment within primary care settings. He has been the recipient of numerous early-career awards, recently including an American Headache Society/Merck U.S. Human Health Scholarship, and he publishes actively in the areas mentioned above.

Chris Spatz is at Hendrix College in Conway, Arkansas. Spatz's undergraduate textbook, *Basic Statistics: Tales of Distributions,* is in its 9th edition (2008). He is also a coauthor with Edward P. Kardas of *Research Methods in Psychology: Ideas, Techniques, and Reports* (2008). Spatz was a section editor for the *Encyclopedia of Statistics in Behavioral Science* (2005) and has reviewed manuscripts for *Teaching of Psychology* for more than 20 years. He has written other chapters for edited books and served as an institutional research consultant to Hendrix College.

Robert J. Sternberg is dean of the School of Arts and Sciences, professor of psychology, adjunct professor of education, and director of the Center for the Psychology of Abilities, Competencies, and Expertise and of the Center for the Enhancement of Learning and Teaching at Tufts University. He is also Honorary Professor of Psychology at the University of Heidelberg. Sternberg is president of the Eastern Psychological Association and president-elect of the International Association for Cognitive Education and Psychology. He is a member of the Board of Directors of the American Psychological Foundation and of the American Association of Colleges and Universities. Sternberg's PhD is from Stanford, and he holds eight honorary doctorates.

Richard Stevenson is an associate professor of psychology at Macquarie University, Sydney, Australia. His PhD work at the University of Sussex was on food preferences, which led to postdoctoral work on food flavor at the CSIRO (Commonwealth Scientific and Industrial Research Organization, Australia) and then at the University of Sydney. He took up an academic post at Macquarie University in 1998 with the aim of applying what he had learned about flavor to odor perception. This culminated in a book on odor perception with Don Wilson in 2006. He is currently shifting back to flavor.

Anne Stewart is professor of graduate psychology at James Madison University. She has worked to promote the resilience of children and families in projects throughout the world, including Sri Lanka and India following the massive tsunami. Dr. Stewart has designed and implemented grant-funded projects to address the psychosocial problems of land mines in Bosnia, Vietnam, Cambodia, and Mozambique. In the United States, she has served as a consultant and service provider after Hurricane Katrina, the 9/11 attacks, the Virginia Tech University shootings, and other catastrophic events. Her books include *Becoming a Community Counselor* and *Thriving! A Manual for Students in the Helping Professions.* She is the president of the Virginia Play Therapy Association and the recipient of the James Madison University All Together One Award and the College Award for Distinguished Service.

Jeffrey R. Stowell earned his PhD in psychobiology from Ohio State University in 1999. His dissertation research was on brain mechanisms that regulate the cardiovascular response to fear and anxiety. Under the direction of Dr. Janice Kiecolt-Glaser at OSU, he completed a postdoctoral fellowship in which he supervised a large study on marital stress, immune function, and wound healing. He currently teaches honors introductory psychology, biological psychology, learning, and controversial topics in psychology at Eastern Illinois University. He has won numerous awards for his teaching, including an Early Career Teaching Excellence Award given by the Society for the Teaching of Psychology in 2006. His current research focuses on how stress and coping influence endocrine function.

Bradley R. Sturz is a recent graduate of the experimental psychology program at Auburn University. After a time as a postdoctoral teacher and researcher at Villanova University, he is now an assistant professor in the Department of Psychology at Armstrong Atlantic State University. His research focus can be broadly classified as comparative psychology with an emphasis in human and nonhuman spatial learning and cognition.

Elizabeth V. Swenson is professor of psychology at John Carroll University. She holds a PhD from Case Western Reserve University and a JD from Cleveland State University. Dr. Swenson teaches professional ethics, planning for graduate school, and legal issues in psychology to undergraduate students. She is a member of John Carroll University's Institutional Review Board, where she routinely reviews student and faculty research proposals for the safeguards to human participants. Dr. Swenson has been a member of the ethics committees of both the American Psychological Association and the Ohio Psychological Association. She practices law in Cleveland, Ohio, in the area of child protection and advocacy.

Laura Ten Eyck is a researcher with the injury prevention program at Children's Medical Center Dallas. She holds several advanced degrees in psychology, including a PhD in experimental social psychology from Texas Christian University. Dr. Ten Eyck has also published several empirical research articles. Along with her colleagues in the injury prevention program, Dr. Ten Eyck applies principles and theories from several fields of psychology to the problem of unintentional injury in children. The goal of the program is to reduce the incidence of preventable injuries in children through basic and applied scientific research. Current injury prevention projects examine the role of attributional biases, personality, and social learning in the transmission of risky behavior and safety attitudes from parents to their children.

Karin Weis is a research associate in the National Preparedness Leadership Initiative of Harvard University. She received her PhD in psychology in 2005 from the University of Heidelberg in Germany. During her doctoral studies, she spent about one and one half years at Yale doing research on love and hate. After receiving her degree, Weis spent 1 year at the University of Connecticut doing postdoctoral work on a fellowship of the Alexander von Humboldt Foundation in Germany. Her research interests include intergroup relations, terrorism, and crisis leadership. She has coauthored a book on hate and has coedited a book on the psychology of love.

James L. Werth, Jr., received his PhD in counseling psychology from Auburn University in 1995 and his Master of Legal Studies degree from the University of Nebraska-Lincoln in 1999. He was the 1999–2000 American Psychological Association William A. Bailey AIDS Policy Congressional Fellow, working on aging and end-of-life issues in the office of United States Senator Ron Wyden (D-OR). When this chapter was written, he was an associate professor in the Department of Psychology at the University of Akron and the pro bono psychologist for the local HIV services organization, where he provided counseling and consulted with a specialty medical clinic. He is currently professor and director of the PsyD program in counseling psychology at Radford University, in Southwest Virginia. He is on the APA Ad Hoc Committee on End-of-Life Issues and the APA Ad Hoc Committee on Legal Issues. He has authored/coauthored over 75 articles and book chapters, edited/coedited 7 special journal issues, and written/edited 4 books specifically on end-of-life matters and/or HIV disease.

Mary Whitehouse holds a Master of Divinity degree from Campbell University and is currently a student at North Carolina State University pursuing her PhD in psychology. She has an interest in the integration of psychology and theology.

Bernard Whitley received his BS in psychology from Loyola University of Chicago and PhD in social psychology from the University of Pittsburgh. He joined the Ball State faculty in 1984 and became department chair in 2005. He is a fellow of the American Psychological Association and a charter member of the Association for Psychological Science. His research interests include academic integrity and prejudice, and he has published in a number of journals, including *Psychological Bulletin*, *Journal of Personality and Social Psychology,* and *Personality and Social Psychology Bulletin.* His books include *The Psychology of Prejudice and Discrimination* (2006) with Mary Kite; *Principles of Research in Behavioral Science* (2nd ed, 2002); *Academic Dishonesty: An Educator's Guide* (2002) with Patricia Keith-Spiegel; *The Ethics of Teaching: A Casebook* (2002) with Patricia Keith-Spiel, Arno Wittig, David Perkins, and Deborah Balogh; and *Handbook for Conducting Research on Human Sexuality* (2002), edited with Michael Wiederman.

Patrick Williams is a graduate student at Claremont Graduate University, with a concentration in cognitive psychology. A lifelong resident of Southern California, Patrick received his bachelor's degree in psychology from California State Polytechnic University, Pomona. His current research involves visuospatial skills training, gender equality, and standardized testing. In our increasingly complex and technological world, the importance of both visuospatial and critical-thinking skills training cannot be underestimated. He is moving forward as a doctoral student in these areas.

William Douglas Woody is an associate professor of psychological sciences at the University of Northern Colorado. He is deeply invested in research and teaching in psychology and law. He conducts research on jury decision making, including topics such as damage awards and jurors' perceptions of juveniles tried as adults, and he teaches classes in psychology and the law. His teaching awards include the Wilbert J. McKeachie Early Career Teaching Excellence Award from the Society for the Teaching of Psychology, the Colorado State University Alumni Association Best Teacher Award, and the University of Northern Colorado Academic Excellence Award for Teaching Excellence in Undergraduate Education. Additionally, he has been named best professor by the students at two of the three universities where he has taught.

Tracy E. Zinn earned her PhD in industrial/organizational psychology with a minor in experimental psychology from Auburn University in 2002. At Auburn, Tracy was privileged to work closely with one of the pioneers in the field of Organizational Behavior Management, and much of her training at Auburn focused on OBM. Currently, she is an assistant professor in the Department of Psychology at James Madison University in Harrisonburg, Virginia, where she teaches, among others, courses in statistics and research methods, industrial/organizational psychology, and OBM. In addition, she conducts research on effective teaching practices, applying behavioral interventions to improve student learning.

PART I

History of Psychology

1

Psychology Before 1900

Ludy T. Benjamin, Jr.
Texas A&M University

In earliest times, basic forms of psychology were practiced by priests, shamans, wizards, seers, medicine men, sorcerers, and enchanters, all of whom offered some blend of magic, religion, and herbal remedies to ease the physical and mental suffering of their patients. These early efforts evolved into the current fields of medicine, religion, and psychology, with a residual overlap of influence among the three. Psychology continued its gradual evolution until the 19th century, one that saw significant changes in this field.

The specific focus of this chapter will be on five forms of American psychology during this period, with some coverage of their European roots. The first is *popular psychology*, or what might be called the psychology of the people, or public psychology. This is the psychology that emerged from thousands of years of human history, manifesting itself in the 19th century in several forms of pseudoscientific psychological practice such as phrenology, physiognomy, mesmerism, spiritualism, and mental healing.

A second psychology, sometimes labeled *medico-psychology* or *psychiatry*, was practiced by physicians and other caretakers in treating the mentally ill who were housed in what were initially called lunatic asylums or insane asylums. These mental asylums were the principal means of care for America's mentally ill for nearly two centuries, and they were the birthplace of the profession of psychiatry. Twentieth-century battles with psychiatry, especially over the right to deliver psychotherapy services, would play a significant role in the development of the profession of clinical psychology.

A third psychology of the 19th century existed within colleges and universities and was known as *mental philosophy*. This academic psychology owed much to the philosophies of the British empiricists and the Scottish realists, whose ideas created an empirical psychology that focused on topics of human consciousness such as sensation and perception, learning, memory, reasoning, attention, and emotion.

In the second half of the 19th century, work in mental philosophy and in sensory physiology and neurophysiology culminated in the development of an experimental approach to the study of mind, an approach that sought to make the questions of mind investigable by experimental methods. This was the fourth form, *scientific psychology*, which drew heavily on the work in physiology and psychophysics in Germany, eventuating in Wilhelm Wundt's 1879 psychology laboratory at the University of Leipzig.

Finally, the fifth psychology to be discussed is *applied psychology*. In the United States, this applied psychology began in the last decade of the 19th century and was manifested initially in the fields of educational psychology, school psychology, clinical psychology, and industrial psychology. In the 20th century this psychology would achieve its greatest growth, forming a new profession of psychology, dominated by the specialty of clinical psychology, whose practitioners became the principal providers of psychotherapy in the last several decades of the 20th century. These five separate but interrelated psychologies are key to understanding contemporary psychology as a science and as a profession, as well as understanding the public's continued attraction to today's popular psychology (see Benjamin, 2007).

NINETEENTH-CENTURY POPULAR PSYCHOLOGY

Although the science of psychology has existed for more than a century, and a profession based on that science has existed for almost as long, these scientifically based psychologies have never supplanted the popular psychology of the public. Today, there are advertisements for palm readers, seers, psychics, spiritualists, mental healers, and others who offer to tell the future, cure depression, save one's marriage, and in general help their clients achieve health, happiness, and success. Books, magazines, TV "psychologists," videos, motivational seminars, and Internet sites will tell you how to lose weight, get a better job, get along better with your spouse, raise your children, be more optimistic, be more assertive, and make more money. Psychology is everywhere. It is the stuff of television, movies, country music lyrics, plays, talk radio, novels, tabloid newspapers, and the Internet. The public cannot seem to get enough of psychology. It is not the kind of psychology that PhD psychologists would call psychology, but it is the public's psychology, and the public embraces it as both entertainment and a prescription for a better life. In the 19th century, Americans also embraced a popular psychology that promised them similar positive outcomes (see T. H. Leahey & G. E. Leahey, 1983).

Phrenology and Physiognomy

Two of the popular psychologies of the 19th century were based on the physical features of a person's head and on the belief that differences in facial features or the shape of the head determined a person's intelligence, personality, integrity, and so forth. *Phrenology* was the invention of Franz Josef Gall (1758–1828), a German anatomist who believed that different brain regions, responsible for different behavioral and intellectual functions, would grow at different rates, producing bumps and indentations on the skull. By measuring these areas of the skull, Gall reasoned that an individual's abilities and behavioral propensities could be measured. Thus a person with a bump above and in front of the ear would be judged to have a tendency to steal, whereas an individual with a bump above and behind the ear would be especially hostile. An enlargement at the top of the forehead signified a person who would be unusually kind.

Phrenology was especially popular in the United States, promoted largely by the efforts of two brothers, Orson Fowler (1809–1887) and Lorenzo Fowler (1811–1896). They opened clinics in New York, Boston, and Philadelphia; published a phrenological magazine for the public; wrote and published many books on the subject; and trained individuals in the "science" of phrenology who then opened their own clinics in various U.S. cities or traveled the countryside taking their phrenological expertise to small towns and rural areas. "Having your head examined" was big business in the 19th century. Some parents would subject a daughter's suitor to such an examination, to see if he was of good character and a promising provider. Some businesses used phrenologists as personnel offices to test prospective employees. Yet most persons sought the services of a phrenologist for personal counseling (see Sokal, 2001). Phrenologists would not only identify areas of personal strength and weakness but also, in the case of weaknesses, prescribe a program of behavioral change. Some bumps signaled the need for "restraint"; some indentations identified characteristics to be "cultivated" (O. S. Fowler & L. N. Fowler, 1859).

Physiognomy, also called characterology, was the evaluation of a person's character, intellect, and abilities based on facial features. It originated in the 18th century in the work of Johann Lavater (1741–1801) and was promoted in the United States by the Fowler brothers and others. The system emphasized the eyes, nose, chin, and forehead as the principal indicators of character. When a young Charles Darwin interviewed for the naturalist position on board the British research ship *Beagle* in 1831, he was almost rejected for the job because the ship's captain, Robert Fitzroy, a believer in physiognomy, determined that the shape of Darwin's nose indicated he was lazy and lacked determination. Fortunately, Fitzroy changed his mind, and Darwin's work on the *Beagle* expedition proved his initial judgments monumentally wrong.

Physiognomy was also used to justify the identification of criminals (see, for example, the work of Cesare Lombroso, 1911), to perpetuate racial and ethnic stereotypes (Wells, 1866), and in many American businesses during the early 20th century, to select employees (see Blackford & Newcomb, 1914).

Mesmerism

In 1775 Franz Anton Mesmer (1734–1815), an Austrian physician, began experimenting with magnets as a means to cure physical and psychological symptoms in his patients. Holding magnets in his hands and moving them over a patient's body, Mesmer reported that his patients experienced a brief fainting spell, after which the problematic symptoms were reduced or eliminated. Mesmer practiced in a time when humoral theories dominated medicine—that is, that bodily fluids or humors such as blood and bile were the keys to good health. Mesmer believed that the body's fluids were magnetized, and that use of the magnets allowed him to "realign" those fluids and thus restore health. Soon Mesmer found that he could dispense with the magnets and merely pass his hands over a patient's body, achieving similar recuperative effects. He reasoned that constant use of the magnets had transferred the magnetic properties to his own hands. Mesmer called his procedure animal magnetism, but as other practitioners began to use the techniques the practice was labeled *mesmerism*.

Mesmer became famous and wealthy, accumulating a clientele that drew heavily from the upper crust of Parisian society. When many individuals, with no medical training, began practicing as mesmerists, it infuriated the French medical community. Physicians asked King Louis XVI to conduct an investigation, which he did, appointing a blue-ribbon commission presided over by Benjamin Franklin. The commission found no evidence that magnetic fluids existed in the body nor that magnetic forces resulted in any healing. No formal actions resulted from the report, and Mesmer and his followers continued their practice.

Mesmerism arrived in the United States in the 1830s and spread, especially in the Northeast. In addition to providing medical healing, mesmerists reportedly cured depression and phobias and helped individuals with self-improvement. Mesmerism is considered to be a precursor of modern hypnotism, a legitimate, if not wholly understood, adjunct to medical and psychological practice. Historians believe that the benefits of mesmerism were the results of suggestions made during hypnotic states. The popularity of mesmerism continued into the early part of the 20th century. It is often regarded as the beginnings of psychotherapy in America (Cushman, 1995).

MEDICO-PSYCHOLOGY OR MENTAL ASYLUM CARE

Large asylums for the care of mentally ill persons were commonplace in Europe in the late 1700s, but by 1800 there were only three such institutions in a more agrarian United States, and all were small, intended to have no more than 250 patients. The asylum movement in America has been described as a "story of good intentions gone bad" (Shorter, 1997, p. 3). Indeed, that phrase is appropriate over and over again in the history of asylums, from the growth of chronic cases in the 19th century that led to the asylums being warehouses of humanity with little or no treatment, to the use of somatic treatments in the early 20th century, such as the prefrontal lobotomy that produced some reduction of psychotic symptoms at a terrible intellectual and emotional cost, to the deinstitutionalization mandated by the Mental Health Act of 1963 that was to produce a more effective community-based mental health treatment program but instead led to thousands of individuals in need of psychological services living on the streets as homeless persons. But at the beginning of the 19th century, there was optimism that these small hospitals for "insane persons" would provide cures in most cases, especially if patients got treatment early.

Asylums were typically constructed in the countryside several miles outside the nearest town. The rural setting removed the patients from the commotion of city life, providing a more pastoral setting. Further, the rural location meant that most asylums had farms where they could grow their own produce and raise their own animals. These farms not only provided food for the asylum kitchen but also provided outdoor work for the asylum inmates, a kind of occupational therapy. Funding for these asylums was never adequate, and asylums fortunate enough to produce surpluses from their farms could sell those crops or animals and use the funds for other needs.

A physician-superintendent, who was often the only physician on staff, headed these asylums. The remaining treatment staff typically consisted of a nurse and minimally trained individuals who served as ward attendants. Treatment in these hospitals in the first half of the 19th century was often referred to as moral therapy, a program intended to return the patients to a productive life in society. Moral therapy included occupational therapy, exercise and recreation, religious training, instruction in good hygiene, and participation in a variety of activities, often selected for patient interests, such as painting, carpentry, gardening, writing, and music. Medical treatments of the day such as bloodletting, cold baths, laxatives, and opium were also used.

As mental asylums grew in America and problems of successful treatment mounted, the superintendents recognized their need for communication with one another. In 1844 they established the Association of Medical Superintendents of American Institutions for the Insane and founded a journal, titled the *American Journal of Insanity*, to publish articles about their asylum work. In 1892 they changed the name of their organization to the American Medico-Psychological Association. In 1918 the name was changed once more to the American Psychiatric Association, and the journal name was changed to the *American Journal of Psychiatry*.

By the middle of the 19th century, much of the optimism about the effectiveness of the asylums had disappeared. Too many of those admitted were not cured, and many had to be housed for the remainder of their lives. In 1869 the Willard State Hospital opened on the shore of Seneca Lake in New York, the first mental asylum built expressly for the chronically insane. By 1875 there were a thousand patients crowded into this asylum of no hope (Grob, 1994). At Willard, and at other asylums where the patient populations had far outgrown the facilities and the staffs, treatment effectively ceased. Staff efforts were directed primarily at patient management. At the end of the 19th century, the patient population in American asylums underwent another period of rapid growth as these asylums were increasingly used as homes for the elderly, especially those suffering from dementia. Although their length of stay typically lasted no more than five years, their numbers were large, and they swelled a system already bursting at the seams. These asylums would grow in number to more than 300 by 1963, when the Mental Health Act brought about the almost immediate release of hundreds of thousands of mental patients. Some of these enormous state asylums (called state hospitals in the 20th century) had populations of more than 7,000 patients. The squalor, hopelessness, sometime cruelty, and general failure of these institutions were documented in 20th-century books and movies. These exposés, which brought public attention to the failures and the development of psychotropic medications in the 1950s, led to the end of the large state hospitals. The eventual failure of the Community Mental Health Center movement of the 1960s was yet another story of "good intentions gone bad."

At the end of the 19th century, some psychologists began to show an interest in the problems of mental illness, a subject that was not in the mainstream of scientific psychology at the time. Several psychologists were hired to work in the mental asylums, principally as researchers. Those positions expanded in the 20th century with more and more psychologists taking on treatment roles. That work would prove formative in the development of the profession of clinical psychology (see Cautin, 2006; Popplestone & McPherson, 1984).

MENTAL PHILOSOPHY

In reading many history of psychology textbooks, it is easy to form the impression that psychology did not exist in American colleges and universities until the science of psychology arrived in the 1880s. Yet there was already an academic psychology in place, known as *mental philosophy*. It was part of departments of philosophy, and it was in these same philosophy departments that scientific psychologists found their first jobs. If you examine the chapter titles in a mental philosophy textbook of the mid-19th century, you would recognize it as a clear precursor to scientific psychology. For example, Thomas Upham's (1848) *Elements of Mental Philosophy*, the most popular mental philosophy textbook of its time, included chapters on sensation and perception, attention, dreaming, consciousness, association (learning), memory, reasoning, emotion, and appetites (motivation). Seven of the first eight chapters of the book addressed sensation and perception, which demonstrated the continued influence of *British empiricism*.

Origin of the empiricist view in philosophy is generally attributed to John Locke (1632–1704). In his most famous work, *An Essay Concerning Human Understanding*, published in 1690, he wrote,

> Let us then suppose the mind to be, as we say, white paper, void of all characters, without any ideas; how comes it to be furnished? Whence comes it by that vast store, which the busy and boundless fancy of man has painted on it with an almost endless variety? Whence has it all the materials of reason and knowledge? To this I answer, in one word, From experience. (Locke, 1690/1849, p. 75)

You will recognize this as the concept of *tabula rasa*, or blank slate, which Aristotle had described centuries earlier and which Locke resurrected. In this monumentally important passage, Locke rejected the notion of innate ideas and asserted that all knowledge, all the contents of the mind, comes from experience. The initial step in that process, he emphasized, was the role of the senses—all knowledge initially comes to the mind from the senses and then from reflection within the mind.

This emphasis on sensation and perception dominated mental philosophy and would play a significant role in the early years of scientific psychology, notably the work of W. Wundt's laboratory in Germany and E. B. Titchener's laboratory at Cornell University in the United States. Empiricism, which began in the late 1600s with Locke, was developed further in the philosophies of George Berkeley, David Hume, James Mill, and John Stuart Mill (1806–1873). In Mill's 1843 book, *A System of Logic*, he argued that the time had come for an empirical science of psychology. Drawing from recent work in chemistry, Mill proposed an empirical psychology focused on sensation that would analyze consciousness into its basic elements, a kind of mental chemistry. His arguments would influence Wilhelm Wundt, who, 36 years later, would move beyond an empirical science to establish psychology as an experimental science.

Also of great importance to mental philosophy was *Scottish realism*, a philosophy that arose in opposition to the British empiricist belief that objects and events in the world were not directly knowable but were the products of sensation and reasoning. Thomas Reid (1710–1796), the founder of Scottish realism, disagreed, arguing instead that perception of worldly events and objects is immediate, and that awareness of such does not require reasoning. This philosophy strengthened the regard for observation as key to a science of psychology. Reid wrote a series of books that focused on the five human senses, describing in detail how the senses are used to acquire knowledge of the world. In two of his books he described the mind in terms of its powers, or faculties, such as perception, memory, and judgment. Reid's taxonomy of mind grew in popularity, leading to the label "Scottish faculty psychology." His ideas crossed the Atlantic with Scottish immigration to the United States and Canada in the late 1700s and early 1800s, and supplanted the ideas of the British empiricists. Evans (1984) has written that by the 1820s, Scottish philosophy dominated American college classrooms.

We have already mentioned Thomas Upham (1799–1872), one of the principal voices in mental philosophy in America. His textbooks dominated the teaching of mental philosophy for more than 50 years. Upham was professor of mental and moral philosophy at Bowdoin College in Maine. A yearlong course in mental and moral philosophy (one semester of each) was common in American colleges in the 19th century. Moral philosophy covered such subjects as morality, conscience, justice, religion, love, and civic duty.

Upham published *Elements of Intellectual Philosophy*, which some historians regard as the first textbook in

American psychology in 1821. He expanded that work to two volumes in 1831 and also wrote a book titled *Elements of Mental Philosophy*. Upham divided mental philosophy into three realms, reflecting the influence of Scottish faculty psychology: intellect, sensibilities, and will. The first volume comprised the intellect and consisted of such topics as the senses, attention, dreaming, consciousness, learning, memory, reasoning, and imagination. The second volume comprised the sensibilities (emotions, desires, moral sensibilities, and abnormal actions and feelings) and the will (Upham, 1831). The work of the mental philosophers in the 19th century was to bring their observational science to bear on those topics, particularly those of the intellect. This work proved foundational for the coming experimental science of psychology. Historian of psychology Alfred Fuchs (2000) has written that "Although the experimental psychologists thought of themselves as replacing a philosophical discipline with a science, they could be more accurately characterized as adding laboratory experimental procedures to what was already defined as an empirical, inductive science" (p. 9).

A SCIENTIFIC PSYCHOLOGY

The 19th century was a period of immense progress in the quest to understand the brain and nervous system. Central to the establishment of psychology as a science was work in neurophysiology and psychophysics.

Neurophysiology

Physiological research involved both structure and function and was focused on questions of nerve specificity, the localization of function in specific regions of the brain, the operations of the senses, and the speed of conductance in the nerves. It was known that the nerves carried information from the senses to the brain (afferent processes) and from the brain to the muscles and effectors (efferent processes). In the first part of the century researchers discovered that nerves in the spinal cord were arranged such that the dorsal part of the cord carried sensory (afferent) information and the ventral part of the cord carried motor (efferent) information. This neural specificity was named the *Bell-Magendie law* after its two discoverers, Charles Bell and François Magendie.

About the same time, in 1826, Johannes Müller (1801–1858) found that sensory nerves carry only one kind of information, regardless of how they are stimulated—a discovery known as the *doctrine of specific nerve energies*. Thus the optic nerve carries only visual information, the auditory nerve carries only sound information, and so forth. This finding gave further support to the idea of organized specificity within the central nervous system. There were those researchers who searched for specificity of function within the brain, believing that specific behaviors were governed by specific brain areas, an idea known as *cortical localization of function*. Paul Broca (1824–1880) was one of the pioneers in this area, locating an area of the left frontal lobe that he found was key to human speech. Today that brain region is known as *Broca's area*.

One of the key beliefs in early 19th-century neurophysiology was that nerves conducted information virtually instantaneously. Müller had written that nerve conductance was so fast that it might occur at the speed of light. Yet by the middle of the 19th century, this belief was being questioned. Hermann von Helmholtz (1821–1894) sought to test *nerve conductance speed* using the severed leg of a frog. He electrically stimulated the upper end of the leg and measured the time required before the foot twitched. The study was a difficult one because of the small time interval involved, but what Helmholtz discovered was quite astounding. He calculated that the impulse traveled at a speed of approximately 90 feet per second. Hardly the speed of light! This discovery would prove key to many studies in the early years of scientific psychology because it was the basis of the reaction-time method.

Psychophysics

October 22, 1850, was a Tuesday. Gustav Fechner (1801–1889) would remember the date exactly because of an incredible insight that occurred to him that day. What was this monumental insight? Fechner realized that it was possible to measure, with great precision, the relation between the physical and psychological worlds—in short, he had discovered a scientific approach to the centuries-old mind-body problem.

The psychological world and the physical world are not the same. If they were, there would be no psychology. All psychological phenomena could be explained fully by the laws of physics. But the physical world and the psychological world are not in a one-to-one correspondence. A pound of feathers and a pound of lead weigh the same in physical terms but when each is placed separately on the outstretched palms of a human subject, the person reports that the lead is much heavier than the feathers. Lights can be made physically brighter that are not perceived as brighter by human observers. Objects can be seen as moving that are actually stationary (one sees motion in motion pictures even though no movement occurs on the screen, only in the viewer's head). Fechner discovered a way to measure the psychological perception of physical events as a way to compare the physical and psychological worlds. Consider his work in difference thresholds, or what is called a *just noticeable difference* or *jnd*. The jnd is the smallest difference that can be detected when a stimulus value is changed. For example, for a weight of 30 ounces, the jnd is 1 ounce. That means a person can reliably differentiate between a weight of 30 ounces and 31 ounces. This relation is lawful, such that the ratio of the change to the value of the stimulus is a constant. So for a 60-ounce weight, the jnd would be not 1 ounce but 2 ounces. Quantitatively, those two jnds are different; once is twice the weight of the other. Yet qualitatively (and this was the key part of Fechner's insight), they represent the same difference! Fechner developed several psychophysical methods for determining difference thresholds and absolute thresholds (the minimal stimulus value that could be detected). These methods were part of the mainstay of the early psychological laboratories and are still used today in psychological research.

Wilhelm Wundt as Founder

Following graduation from medical school at the top of his class in 1855, Wilhelm Wundt (1832–1920) worked as a research assistant, first for Johannes Müller and then Hermann Helmholtz. He was well versed in the physiology and psychophysics of the 19th century and was particularly influenced by Fechner's 1860 book, *Elements of Psychophysics*. Wundt's vision for a scientific psychology gestated for nearly 20 years. In 1874 he published the first textbook of the scientific work relevant to psychology, *Principles of Physiological Psychology*. In that book he left no doubt about his intent: "The book which I here present to the public is an attempt to mark out a new domain of science" (Wundt, 1904, p. v). Five years later, in 1879 at the University of Leipzig, he established the world's first psychological laboratory. He would remain active in his laboratory into the 1910s, mentoring more than 160 doctoral students in psychology and philosophy.

Much of the research in Wundt's laboratory was on sensation and perception, reflecting the influence of the British empiricists. Soon the volume of research required a journal to publish the results, so in 1881, Wundt began publication of *Philosophical Studies*, the first journal for scientific psychology (the title *Psychological Studies* was already in use for a parapsychology journal and thus could not be used). Wundt also drew on Helmholtz's work on the speed of nerve conduction as a basis for studying the time course of mental events.

Using a reaction time technique pioneered by the Dutch physiologist F. C. Donders, Wundt measured the difference between a simple reaction time (recognizing when a light was presented) and a choice reaction time (recognizing whether the light was green or red and then responding to each one differently). The longer reaction times in the second task were seen to represent the processing times of the mental events in the complex task that were not involved in the simple task. Thus by further complicating various reaction time tasks, it was possible to measure the speed of thinking.

For Wundt, the goal of psychology was to understand consciousness, which Wundt defined as the sum total of all facts of which a person is conscious. Consciousness was built up through an individual's experiences in the world. These experiences, according to Wundt, were composed of sensations, associations, and feelings. These were the elements of consciousness, and Wundt proposed studying those elements to discover the ways in which they were combined to form what he called psychical compounds or aggregates. The compounds were of paramount importance. Wundt (1902) wrote, "The actual contents of psychical experience…[depend] for the most part, not on the nature of the elements, so much as on their union into a composite psychical compound" (p. 33). Thus, the mind was an active entity that organized, analyzed, and altered the psychical elements and compounds of consciousness, creating experiences, feelings, and ideas that were not evident in any study of each component by itself. Wundt called this psychological system *voluntarism* to indicate the voluntary, active, and willful nature of the mind.

Wundt's laboratory was soon joined by other psychology laboratories in Germany, including those of Hermann Ebbinghaus at the University of Berlin and Georg Müller at Göttingen in the 1880s, both of whom did important work on memory, and Carl Stumpf at Berlin and Oswald Külpe at the University of Würzburg in the 1890s. Wundt also trained many of the American psychologists who would return to found laboratories in their own country—for example, James McKeen Cattell at the University of Pennsylvania (1889), Harry Kirke Wolfe at the University of Nebraska (1889), Edward Pace at Catholic University (1891), Frank Angell at Cornell University (1891), Edward Scripture at Yale University (1892), George Stratton at the University of California-Berkeley (1896), and Walter Dill Scott at Northwestern University (1900). By 1900 there were 40 psychology laboratories in the United States (Benjamin, 2000), and many others had spread across Europe as well.

Scientific Psychology in the United States

The United States generated its own homegrown science of psychology, principally from the influence of William James (1842–1910) and his student G. Stanley Hall (1844–1924). James began teaching a course on scientific psychology at Harvard University in 1875, and Hall would earn his doctorate with James there in 1878. James was not a laboratory psychologist, although he did have a laboratory for demonstration purposes. James's chief influence on psychology—and it was an incredibly significant one—was his writing of the *Principles of Psychology*, published in 1890, which many historians consider the single most important psychological work in the English language. James's two-volume work influenced a generation of students to pursue this new science. Although James's book was never revised, it remains in print today and has been continuously available since its publication. Whereas James did not revise his 1400-page magnum opus, he did write a brief edition in 1892 titled *Psychology (Briefer Course)*, which served more commonly as a college textbook.

Upon graduation, Hall pursued additional study in Germany in 1879 and, for a brief period, worked in Wundt's laboratory in its initial year. Eventually Hall would establish his own psychology laboratory in 1883 at Johns Hopkins University, the first such laboratory in America. His students would soon found other psychology laboratories at Indiana University, the University of Wisconsin, Clark University, and the University of Iowa. Hall was an exceptionally ambitious man (see Ross, 1972), and he did much to promote the development of the new science. He founded the first psychological journal in America, the *American Journal of Psychology*, first published in 1887. He founded the American Psychological Association (APA) in 1892 and served as its first president. In the 1890s he was an important force in launching the child study movement, which will be discussed later. His research focused mostly on what would today be called developmental psychology. His most important contribution in that regard was a two-volume work titled *Adolescence: Its psychology and its relations to physiology, anthropology, sociology, sex, crime, and religion* (1904). This work was important, in part, because it changed the word *adolescence* from a largely technical term to a word in everyday usage in the English language.

In the year that Hall founded the APA, significant happenings were occurring in the laboratories of American psychology. Four of Wundt's doctoral students, two of them American, came to the United States that year to serve as laboratory directors at their respective universities. The Americans were Lightner Witmer, who took over Cattell's lab at the University of Pennsylvania after Cattell moved to Columbia University in 1891, and Edward Wheeler Scripture, who was in charge of the laboratory at Yale. Edward B. Titchener, an Englishman, arrived that same year to direct psychology at Cornell University, and Hugo Münsterberg, a German, was recruited to Harvard University by William James to direct that laboratory. Each of these individuals would prove especially important to the development of American psychology.

Structuralism

Edward Bradford Titchener (1867–1927) was 25 years old when he arrived at Cornell University. He established a brand of psychology there that would become known as *structuralism*, a reductionistic approach to the study of consciousness that sought to study consciousness by breaking it down into its fundamental elements, namely sensations, feelings, and images. Identifying the elements of consciousness was the first of his goals for psychology. A second was to discover how the elements became grouped and arranged, and a third was to determine the causes of the particular arrangements of elements. Titchener believed that the causes of these elemental arrangements would be found in underlying physiological processes.

For his research, Titchener relied mostly on a single method—*introspection*, or looking within. (See the following chapter for additional information regarding Titchener and introspection.) Titchener put his introspectors (his graduate students) through rigorous training to develop in them what he called the introspective habit. In introspection the individual was to report on mental events, describing them as accurately and completely as possible, striving to report one's consciousness at its most basic level, that is, in terms of its elements. Training in introspection allowed observers to function automatically in making internal observations, with minimal disruption of the mental event being studied. Titchener's introspectors were expected to avoid making the *stimulus error*, that is, confusing what was being observed (e.g., a book) with the basic elements of that stimulus (e.g., color, texture, shape). It was these elements that Titchener believed were the fundamentals of conscious experiences, and he expected the introspective accounts in his laboratory to be at that level. Titchener's research focused on sensations that he studied in terms of their primary attributes: quality (cold, red, salty), intensity (bright, loud), clearness (distinct vs. indistinct), and duration (time course of the sensation). He and his students identified more than 44,000 separate, distinct sensory experiences (elements) in vision and audition alone (see Evans, 1984; Titchener, 1898)

Whereas Titchener's psychology focused on the structure of consciousness (its elemental content), a rival psychological school—*functionalism*—was more interested in the functions of consciousness, that is, what is it that consciousness does for the organism? Why does consciousness exist? It might be obvious that the ideas of Charles Darwin, especially the issues of individual differences and adaptation, would be of special interest to this psychological group.

Functionalism

American functional psychology was rooted in the ideas of William James, G. Stanley Hall, and James McKeen Cattell. James argued that consciousness enabled the individual to make choices, which was its adaptive significance. Consciousness helped an organism to maintain attention in situations of stimulus overload, to attend selectively when selection was an advantage, and to make use of habitual responding when such automaticity was adaptive. Hall and Cattell were particularly interested in applying the science of psychology—Hall to education and Cattell to mental testing, especially intelligence testing. A strong interest in the applications of scientific psychology was a key part of functionalism, whereas Titchener was adamant that application was premature until the structure of consciousness was completely worked out. Addressing the functionalists' interests in applied work, Titchener (1910) wrote, "The diversion into practical channels of energy which would otherwise have been expended in the service of the laboratory must be regarded as a definite loss to pure science" (p. 407).

Whereas structuralism existed principally at Cornell University, the University of Chicago served as the stronghold for functionalism. Head of the psychology department there at the end of the 19th century was James Rowland

Angell (1869–1949), who had studied with William James and John Dewey before pursuing further graduate work in Germany. (See the following chapter for additional information on James Rowland Angell and functionalism.)

The structuralists had emphasized the study of sensations; the functionalists would emphasize the study of learning because it was through learning that consciousness could serve its adaptive function. In their research the functionalists used introspection, but not in the way Titchener advocated. Instead, their method was more of a self-report of experience, usually of discreet, brief sensory events. They also used psychophysical methods, questionnaires (made popular in the work of Hall), mental tests (many of them devised by Cattell), and animal studies in a program of comparative psychology in which nonhuman organisms were studied in order to generalize findings to human behavior. In addition to learning they studied sensation and perception, child development, intelligence, sex differences, motivation, abnormal behavior, personality, and other topics. Further, they also used their science to investigate applied questions in education, business, and the law. Both schools of psychology would lose their influence in the 1920s with the rising dominance of behaviorism in American psychology (O'Donnell, 1985).

THE NEW APPLIED PSYCHOLOGY

In contemporary psychology, applied psychology dominates the field, principally in the practice of clinical psychology but in many other applied specialties as well. Yet when scientific psychology began in the last quarter of the 19th century, application of the new science was not an immediate goal. The new psychologists sought to understand the normal mind, to explore ordinary mental life in terms of sensation, perception, learning, thinking, and so forth. As we have described earlier, when scientific psychology arrived on the scene, an applied psychology already existed. It could be found in the field of psychiatry as it labored to cure persons with mental illnesses, and it could be found in popular psychology in the many forms of pseudoscientific practice such as phrenology and mesmerism. A new applied psychology, however, would develop from the scientific laboratories, promoted largely by those identified with functional psychology. The earliest of these applications was in the field of education.

Child Study

At the end of the 19th century, America was undergoing significant social change. Cities were experiencing expansive growth brought on by the industrial revolution. Exploitation of children in factory jobs led to new child labor laws and compulsory school attendance laws that forced children into longer stays in school. New waves of immigrants were arriving in America, also swelling the enrollments in schools. Concerns about juvenile delinquency, about integrating new immigrants into American society, and about the explosive growth of school enrollment brought a new focus to the American educational system and new demands for reform. In the early 1890s, G. Stanley Hall called for a national program of what was labeled "child study." Psychologists, educators, and parents were to band together in a program of research to learn all there was to know about children. Everything was to be studied: sensory capabilities, religious beliefs, physical characteristics, sense of humor, memory, play, language development, personality, attention span, and so forth. Hall's belief was that armed with this new knowledge of the child, education could be vastly improved through better teacher training and more individualized and scientifically designed school curricula. In a popular magazine article written in 1894, Hall touted the applicability of his new science: "The one chief and immediate application for all this work is its application to education" (p. 718). The child study movement, which lasted until about 1905, never achieved the lofty goals that Hall and others imagined (see Davidson & Benjamin, 1987; Ross, 1972). Although child study proved to be more popular psychology than scientific psychology, no theory guided the research, and it proved impossible to integrate the disparate results of the many questionnaire studies into any meaningful prescriptions for educational change. It was, however, the first principal application of the new scientific psychology, and it did foster the development of several fields of psychology that would prove important, notably developmental psychology and educational psychology.

Clinical and School Psychology

In March 1896, a schoolteacher visited Lightner Witmer (1867–1956) at the psychology department of the University of Pennsylvania. The teacher brought with her a 14-year-old boy who had difficulty spelling. The teacher believed that the problem was a mental one, and she reasoned that if psychology was the science of the mind, then psychology ought to be able to help. Witmer accepted the challenge and was able to develop a treatment program that helped the boy considerably. As word spread about his success with this case, other children with educational and behavioral problems were brought to him. Thus was born the first psychological clinic in America. By December of that year, Witmer was so pleased with the outcomes of the work in his clinic that he gave an address promoting applied psychology at the annual meeting of the American Psychological Association in which he urged his colleagues to use psychology "to throw light upon the problems that confront humanity" (Witmer, 1897, p. 116).

Soon the Philadelphia public schools began to use Witmer's clinic on a routine basis, referring children there who seemed to be in need of the kinds of diagnostic and therapeutic services offered. The growth of the clinic led Witmer in 1907 to establish a journal, titled *The Psychological Clinic*, to publish descriptions of the more important cases that had been treated successfully. This work proved to be an important precursor in the fields of clinical psychology and school psychology, and it is for that reason that Witmer is generally labeled the founder of both fields (see Baker, 1988; McReynolds, 1997).

Another important precursor of clinical psychology can be found in the 1890s work of James McKeen Cattell (1860–1944) in the field of mental testing. Cattell (1890) coined the term "mental test" and developed numerous tests that assessed sensory, motor, and cognitive functioning. His tests were considered to be measures of intelligence until research at the beginning of the 20th century showed poor correlations between his tests and academic performance. His intelligence tests were supplanted by tests developed by French psychologist Alfred Binet (1857–1911). Intellectual assessment, primarily using the Binet tests, would become the most characteristic activity of clinical psychologists in the first half of the 20th century (Sokal, 1987).

The Psychology of Business

By 1900 scientific psychology had not only moved into the domains of clinical and educational work but had also initiated work in the field of business. The initial foray was in the field of advertising. According to historians of American business Bryant and Dethloff (1990), "Advertising became increasingly important after the 1890s as manufacturers reduced price competition and stressed product differentiation [and as] manufacturing capacity came to exceed demand…In an economy of excess, advertising became the means to dispose of an oversupply" (p. 190). Companies greatly expanded their advertising budgets to deal with product surpluses, and they grew increasingly concerned about maximizing the effects of their advertising campaigns. Because advertising was about appeal and persuasion of customers, businesses looked to the new science of psychology for help. In 1895 the University of Minnesota's Harlow Gale (1862–1945), who had studied with Wundt, conducted a survey of approximately 20 businesses, asking about their advertising strategies. His work, although the first of the advertising studies in psychology, had little, if any, impact on American business. But Walter Dill Scott (1869–1955), a psychologist and a Wundt doctoral student, made a significant contribution to the field, changing the way businesses advertised, leading them to create advertisements that had greater suggestive appeal. Scott argued that humans were nonrational, highly suggestible creatures "under the hypnotic influence of the advertising writer" (Kuna, 1976, p. 353). Scott advocated two advertising methods in particular: the direct command and the return coupon. Both techniques relied on Scott's belief that individuals were compulsively obedient. Scott's early work in this field appeared as a series of monthly articles in advertising magazines. These were later collected in the first two books on the psychology of advertising (Scott, 1903, 1908), the principal beginnings of the field known today as industrial-organizational psychology.

SUMMARY

The 20th century would soon prove to be one of enormous progress in the discipline of psychology, witnessing the division of the field into numerous scientific and practice specialties. Specialties such as clinical psychology, school psychology, industrial-organizational psychology, and counseling psychology dominated the applied fields, and newer applied specialties such as health psychology, forensic psychology, and sport psychology would appear near the end of the century (see Benjamin & Baker, 2004).

The "schools" of American psychology, manifested in structuralism and functionalism in the 19th century and in behaviorism for much of the 20th century, gave way to divisions of scientific psychology by subject matter, such as social psychology, cognitive psychology, developmental psychology, and behavioral neuroscience. In examining the continuing evolution of psychology in the 21st century, both as a science and as a field of professional practice, it is important to recognize that the seeds for these areas were well established in the fertile soil of the 1800s.

REFERENCES AND FURTHER READINGS

Angell, J. R. (1906). The province of functional psychology. *Psychological Review, 14,* 61–91.

Baker, D. B. (1988). The psychology of Lightner Witmer. *Professional School Psychology, 3,* 109–121.

Benjamin, L. T., Jr. (2000). The psychology laboratory at the turn of the 20th century. *American Psychologist, 55,* 318–321.

Benjamin, L. T., Jr. (2007). *A brief history of modern psychology*. Malden, MA: Blackwell Publishing.

Benjamin, L. T., Jr., & Baker, D. B. (2004). *From séance to science: A history of the profession of psychology in America*. Belmont, CA: Wadsworth.

Blackford, K. M. H., & Newcomb, A. (1914). *The job, the man, the boss*. New York: Doubleday & Page.

Bryant, K. L., Jr., & Dethloff, H. C. (1990). *A history of American business* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15,* 373–381.

Cautin, R. L. (2006). David Shakow: Architect of modern clinical psychology. In D. A. Dewsbury, L. T. Benjamin, Jr., & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 6, pp. 207–221). Washington, DC: American Psychological Association.

Cushman, P. (1995). *Constructing the self, constructing America: A cultural history of psychotherapy*. Reading, MA: Addison-Wesley.

Davidson, E. S., & Benjamin, L. T., Jr. (1987). A history of the child study movement in America. In J. A. Glover & R. R.

Ronning (Eds.), *Historical foundations of educational psychology* (pp. 41–60). New York: Plenum Press.

Evans, R. B. (1984). The origins of American academic psychology. In. J. Brozek (Ed.), *Explorations in the history of psychology in the United States* (pp. 17–60). Lewisburg, PA: Bucknell University Press.

Evans, R. B. (1985). E. B. Titchener and American experimental psychology. In H. Carpintero & J. M. Peiro (Eds.), *Psychology in its historical context* (pp. 117–125). Valencia, Spain: Universidad de Valencia.

Fechner, G. T. (1966). *Elements of psychophysics* (H. E. Adler, Trans.). New York: Holt, Rinehart and Winston. (Original work published 1860)

Fowler, O. S., & Fowler, L. N. (1859). *New illustrated self-instructor in phrenology and physiology.* New York: Fowler and Wells.

Fuchs, A. H. (2000). Contributions of American mental philosophers to psychology in the United States. *History of Psychology, 3,* 1–18.

Grob, G. N. (1994). *The mad among us: A history of the care of America's mentally ill*. Cambridge, MA: Harvard University Press.

Hall, G. S. (1894, August). The new psychology as a basis of education. *Forum*, pp. 710–720.

Hall, G. S. (1904). *Adolescence: Its psychology and its relations to physiology, anthropology, sociology, sex, crime, and religion* (2 vols.). New York: D. Appleton and Co.

James, W. (1890). *The principles of psychology* (2 vols.). New York: Henry Holt.

James, W. (1892). *Psychology (Briefer course)*. New York: Henry Holt.

Kuna, D. P. (1976). The concept of suggestion in the early history of advertising psychology. *Journal of the History of the Behavioral Sciences*, *12,* 347–353.

Leahey, T. H., & Leahey, G. E. (1983). *Psychology's occult doubles: Psychology and the problem of pseudoscience*. Chicago: Nelson Hall.

Locke, J. (1849). *An essay concerning human understanding*. Philadelphia: Kay and Troutman (Original work published 1690)

Lombroso, C. (1911). *Criminal man*. New York: G. Putnam's Sons.

McReynolds, R. (1997). *Lightner Witmer: His life and times*. Washington, DC: American Psychological Association.

Mill, J. S. (1843). *A system of logic, racioinative and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation*. London: John W. Parker.

O'Donnell, J. M. (1985). *The origins of behaviorism: American psychology, 1870–1920*. New York: New York University Press.

Popplestone, J. A., & McPherson, M. W. (1984). Pioneer psychology laboratories in clinical settings. In J. Brozek (Ed.), *Explorations in the history of psychology in the United States* (pp. 196–272). Lewisburg, PA: Bucknell University Press.

Ross, D. (1972). *G. Stanley Hall: The psychologist as prophet*. Chicago: University of Chicago Press.

Scott, W. D. (1903). *The theory of advertising*. Boston: Small, Maynard.

Scott, W. D. (1908). *The psychology of advertising*. Boston: Small, Maynard.

Shorter, E. (1997). *A history of psychiatry: From the era of the asylum to the age of Prozac.* New York: John Wiley.

Sokal, M. M. (Ed.) (1987). *Psychological testing and American society, 1890–1930*. New Brunswick, NJ: Rutgers University Press.

Sokal, M. M. (2001). Practical phrenology as psychological counseling in the 19th-century United States. In C. D. Green, M. Shore, & T. Teo (Eds.), *The transformation of psychology: Influences of 19th-century philosophy, technology, and natural science* (pp. 21–44). Washington, DC: American Psychological Association.

Titchener, E. B. (1898). The postulates of a structural psychology. *Philosophical Review, 7,* 449–465.

Titchener, E. B. (1910). The past decade in experimental psychology. *American Journal of Psychology, 21,* 404–421.

Upham, T. C. (1827). *Elements of intellectual philosophy*. Portland, ME: William Hyde.

Upham, T. C. (1831). *Elements of mental philosophy* (2 vols.). Portland, ME: Hilliard Gray and Co.

Upham, T. C. (1848). *Elements of mental philosophy* (2 vols.). New York: Harper & Bros.

Wells, S. (1866). *New physiognomy or signs of character as manifested through temperament and external forms and especially in the human face divine*. New York: Fowler and Wells.

Witmer, L. (1897). The organization of practical work in psychology. *Psychological Review, 4,* 116–117.

Wundt, W. (1874). *Grundzüge der physiologischen Psychologie* [Principles of physiological psychology]. Leipzig: Wilhelm Englemann.

Wundt, W. (1902). *Outlines of psychology* (C. H. Judd, Trans.). New York: Gustav E. Stechert.

Wundt, W. (1904). *Principles of physiological psychology* (5th ed., E. B. Titchener, Trans.). New York: Macmillan.

2

Psychology in the 20th Century

C. James Goodwin
Western Carolina University

On the final page of his *Psychology (Briefer Course)* (1892), a condensed version of his monumental *Principles of Psychology* (1890), Harvard psychologist/philosopher William James reflected on the "New Psychology" of the 19th century and concluded that psychology "is no science, it is only the hope of a science" (p. 468). In the 20th century that hope became reality. This chapter, necessarily incomplete, will trace the evolution of the discipline of psychology from a potential science at the start of the 20th century to a real one by the end of the century. Any organizational scheme will be arbitrary, and the choice here is topical, with the chapter segmented into two main sections that roughly parallel the distinction between broad categories of activity—research and practice. A brief third section will describe four trends that were evident near the end of the 20th century. Although psychology is an international discipline, the focus of this chapter will be the development of psychology in the United States.

THE RESEARCH SIDE: SCHOOLS OF THOUGHT IN ACADEMIC PSYCHOLOGY

American psychology at the start of the 20th century was the psychology of human consciousness and mental life. There was also strong interest in animal behavior, primarily as it related to the question of the evolution of human mental processes. This new discipline of psychology emerged in the second half of the 19th century, the offspring of recurring philosophical questions about human nature, advances in knowledge about the physiology of the brain and nervous system, and the strong influence of Darwinian evolutionary biology. This so-called "New Psychology," in contrast to the older psychology of philosophical speculation and logical analysis, was one that developed in an academic environment, through the creation of research laboratories for the study of such cognitive phenomena as sensation, perception, association, attention, and memory. By 1900, there were no fewer than 41 laboratories of experimental psychology in America (Benjamin, 2000), ranging from the well established (e.g., Cornell University, Clark University) to the fledgling (e.g., University of Maine, Northwestern University).

Structuralism and Functionalism

Edward B. Titchnener

In the first decade of the 20th century, one major issue that occupied psychologists in America concerned the primary goal for this new laboratory psychology. The question divided advocates of two schools of thought who have come to be known as structuralists and functionalists, and led directly to the development of a third school, behaviorism. The central figure of the structuralist school was E. B. Titchener (1867–1927), a British psychologist with a German temperament who earned a PhD with

Wilhelm Wundt at Leipzig in 1892 and in that same year came to Cornell University, where he spent his entire career. For Titchener, the new psychology was to be the study of the basic structure of adult human consciousness, to be studied in the laboratory using precise instruments (e.g., chronoscopes for measuring the amount of time it took to respond to various stimuli) and a self-report procedure called *systematic experimental introspection*. To introspect in Titchener's laboratory meant to participate in some lab experiment (e.g., a reaction time task) and then to give a detailed verbal description of the mental processes involved in the experience. Introspectors had to be highly trained to avoid bias. That is, they were to become introspecting machines. The overall purpose of this research was to analyze human conscious experience into its fundamental structural components—Titchener eventually concluded that these basic elements were sensations (the basic element underlying perception), images (the basic element of thought), and the affective states of pleasantness and unpleasantness (the basic element of emotion). Structuralism as a system did not outlive Titchener and has little relevance today, and introspection was eventually discarded as being inherently subjective, but Titchener remains an important figure in psychology's history for one main reason. More than anyone else in psychology's early history, he was a passionate advocate of the value of basic research in psychological science as an acknowledgment of the importance of having a clear understanding of fundamental human mental processes. As one historian wrote, Titchener was "responsible for making psychology scientific—for setting psychology up in the laboratory and reaching his conclusions through experiment, under controlled conditions" (Hindeland, 1971, p. 28).

Although Titchener was a commanding presence, most American psychologists believed that the structuralist approach was at best incomplete, and they became associated with a school of thought called functionalism. This was a way of thinking that had direct ties to Darwinian evolutionary thinking, and it is a way of thinking that permeates American psychology to this day. Functionalists believed that, rather than the study of structure of human consciousness, psychology's goal ought to be to understand how human conscious experience enables the organism to adapt to the environment and thereby prosper. As James Angell put it in "The Province of Functional Psychology," his famous 1906 address as president of the American Psychological Association (APA), structuralists were interested in the question "What is mind?" whereas functionalists wanted to know "What is mind for?" To answer the question of how human consciousness functioned, psychologists were led to investigate a much wider range of topics than those of interest to Titchener. For example, although Titchener considered the study of children to be outside the realm of psychology (children were not considered capable of accurate introspection), functionalists were interested in children to learn something about how the ability to adapt developed in them.

William James

Although he would not have labeled himself a functionalist, William James (1842–1910), whose concerns about psychology as a science opened this chapter, epitomized the spirit of this school of psychology. Today we recognize James as one of psychology's brightest lights, and his *Principles of Psychology* (1890) is considered psychology's most important book. As an indication of the esteem in which his peers held him, he was elected to the presidency of the APA twice, and when James McKeen Cattell of Columbia asked psychologists to rank their peers for eminence in 1903, James was ranked first on *every* return (Hothersall, 2004).

James is perhaps best known for his writings about human consciousness, habit, and emotion. In direct contrast to Titchener's goal of analyzing consciousness into its structural elements, James argued that such an analysis was artificial, destroying the nature of the phenomenon. Instead, he argued, consciousness should be thought of as analogous to a flowing stream—it is a continuous action, never quite the same from moment to moment. Rather than focus on structure, James chose to emphasize the manner in which human consciousness helps the individual function effectively in the world. For James, the essential function of human consciousness was to allow the person to adapt quickly to new challenges found in the environment, to learn new things quickly, and to solve problems efficiently. Habits also had an important and related function. According to James, because habits occurred more or less automatically, with little conscious thought, they enabled the person to focus conscious effort on other, more important problems. He also had some practical suggestions about how to form successful habits, such as making a public announcement of intent (e.g., telling friends of a plan to walk two miles a day).

The Jamesian idea that is most likely to be found in modern psychology textbooks is the James-Lange theory of emotion (Lange was a Dutch physiologist whose ideas were similar to James's). According to this theory, the bodily changes (e.g., the quick intake of breath and sense of one's heart stopping when a snake is spotted on a mountain trail) that accompany emotional experiences are the essence of the emotion itself. Remove the bodily reaction and the emotion is gone. Similarly, simulate bodily actions associated with emotion, and the emotion will appear. The theory failed to account for the complexity of physiological arousal (no fault of James—sophisticated brain research was still to come), but to this day, the theory has a degree of truth to it. For instance, some research shows that by forcing people to use their facial muscles in a way that simulates the facial expression of a particular emotion, that emotion will be experienced.

Other Prominent Pioneers

E. B. Titchener and William James were but two of an extraordinary group of individuals who helped create the new discipline of psychology. Other prominent first- and second-generation American psychologists included G. Stanley Hall, James McKeen Cattell, Mary Whiton Calkins, Margaret Washburn, Robert Woodworth, and Edward Thorndike.

G. Stanley Hall

Hall (1844–1924) is sometimes referred to by historians as “Mr. First.” Among other things, he created the first laboratory of experimental psychology in America (Johns Hopkins in 1883), and the first journal of psychology in America (*American Journal of Psychology* in 1887); he was the first president of Clark University, which opened in 1889; he founded the APA in 1892 and was its first president; and he was largely responsible for bringing Freud’s ideas to America, by inviting the Viennese physician to America in 1909 to celebrate Clark’s 20th birthday. He was also a prominent developmental psychologist, and his comprehensive 1904 text, *Adolescence*, is still considered the book that first identified the teen years as a distinct developmental stage, and that created the modern subdiscipline of adolescent psychology.

James McKeen Cattell

Cattell (1860–1944) earned a doctorate with Wundt at Leipzig in 1886, but Britain’s Sir Francis Galton influenced him more. Galton reinforced Cattell’s beliefs that the study of how individuals differed from one another was as important as the search for general principles that applied to everyone. Hence, Cattell’s focus on individual differences contrasted with Titchener’s goal of discovering general laws of mental life. The interest in how people differed led Cattell to the issue of how these differences could be measured. He coined the term “mental test” in 1890, although the tests he developed (e.g., grip strength, two-point threshold, time to name colors) had more to do with sensory and muscular capacity than with conceptual ability. Consequently, when he correlated scores on his mental tests with college grades, in the hope that such tests might be useful as predictors of college success, the outcome was a failure, leading him to abandon the effort. The bulk of his career concerned the professionalization of psychology—he was active in APA, he founded journals, and, as head of the psychology program at Columbia, he produced a significant number of PhDs who became well known themselves (e.g., Thorndike, Woodworth).

Mary Whiton Calkins

The career of Mary Calkins (1863–1930) of Wellesley College illustrates many of the difficulties faced by women in the late 19th and early 20th centuries. Although she was invited to take classes at Harvard as a guest (on the strong recommendation of William James), she was never granted the PhD that she earned there because Harvard refused to officially accept women as graduate students. Yet, at Harvard, she completed important research on memory and association that featured the invention of what became a standard method in memory studies, paired associates learning. Her research uncovered a number of memory phenomena (e.g., primacy and recency effects) that were later investigated more thoroughly when cognitive psychology became a force in the second half of the 20th century. Her peers recognized her importance when they elected her the first woman president of the APA (14th APA president overall) in 1905.

Margaret Washburn

Unlike Harvard, Cornell University admitted women graduate students, and Margaret Washburn (1871–1939) became the first woman to earn a PhD in psychology in 1894. Although she was Titchener’s first doctoral student, she had little interest in his structuralist school. She completed important research on perception, imagery, and “social consciousness” (empathy and helping behavior), but she is best known as a leading comparative psychologist. She contributed some original research (on color vision in fish, for instance) and summarized the field in a 1908 text, *The Animal Mind*; it became the standard comparative psychology text of its day. She was elected APA president in 1921, co-edited the *American Journal of Psychology* for more than a decade, and was elected to the prestigious National Academy of Sciences in 1931.

Robert Woodworth

Students in research methods courses often struggle with the distinction between independent and dependent variables, and they often confuse experimental and correlational studies, along with the proper conclusions to be drawn from each. For these frustrations, they have Robert Woodworth (1869–1962) to thank. Woodworth was a legendary teacher of methodology at Columbia who, in 1938, published a research textbook with the simple title of *Experimental Psychology* that became known as the “Columbia bible.” It became a standard methodology text, training untold thousands of psychologists in how to do research (Winston, 1990). It was Woodworth who created the modern definition of an experiment—manipulating an independent variable, controlling other factors, and then measuring the outcome or dependent variable. And it was Woodworth who first drew the distinction between the experiment and a correlational study, with the latter measuring the strength of relation between two measured variables, but not directly manipulating an independent variable.

Edward Thorndike

Thorndike (1874–1949), for many years a colleague of Woodworth’s at Columbia University, was a transitional figure between functionalism and behaviorism (below). His doctoral dissertation on animal learning became one of psychology's most famous series of studies. What Thorndike did was to build small "puzzle boxes" with doors that could be opened in various ways (e.g., by pushing a pedal on the floor) by the cats placed in them. The cats would make various attempts to escape the boxes, eventually pushing the pedal. On subsequent attempts they would solve the puzzle box sooner and sooner, demonstrating a trial-and-error learning process (Thorndike called it trial and accidental success). On the basis of this research, Thorndike proposed his "Law of Effect"—in essence, it stated that over a series of trials, successful behaviors would be "stamped in" by the reward of escape, whereas unsuccessful behaviors would be "stamped out" (Thorndike, 1911). B. F. Skinner later credited Thorndike for being the originator of the learning model that came to be known as operant or Skinnerian conditioning.

Gestalt Psychology

As a school of thought, the origins of German Gestalt psychology lie in the late 19th century, as a protest against a structuralist strategy for understanding mental life. In the 1920s and 1930s, several leading gestalists came to the United States, and the focus of their protest shifted to behaviorism (next section). In essence, the gestalists argued that when trying to understand some psychological phenomenon, it was a mistake to try to analyze it into its constituent elements, whether they be the sensations or images of the structuralists or the stimulus-response units of the behaviorists. Rather, the whole of the phenomenon was always more than the sum of its individual elements, according to the gestalt mantra. For example, a song is composed of a sequence of notes, but the song is more than the sum of those individual notes. If the song were sung in a different octave, every note would change, but the song's melody would retain its identity.

The founding of the Gestalt psychology movement is usually attributed to the German psychologist Max Wertheimer (1880–1943); his experiments on apparent motion (he preferred the term *phi phenomenon*) are considered to be the starting point for the gestalt school. The demonstration is a simple one, accomplished by flashing a vertical light for a brief time, then a horizontal light. If the timing is right, the perception is not of two lights, but of a single light falling from a vertical into a horizontal position. Wertheimer argued that it was simply impossible to analyze such an experience into its sensory components because the perception of the light between the vertical and horizontal positions does not have a sensory element to it. The whole (perception of a light falling over) is more than the sum of its individual parts (the vertical light, and then the horizontal light). Wertheimer also wrote extensively about a series of perceptual organizing principles (e.g., closure, figure-ground) that we use to interpret sensory information meaningfully, and later in his life, he wrote on thinking and problem solving.

Another prominent Gestalt psychologist was Wolfgang Köhler (1887–1967), best known for his work on problem solving in apes. This took place on the island of Tenerife, the largest of the Canary Islands, which lie off the northwest coast of Africa. The Prussian Academy of Sciences had a primate colony on the island, and in 1913, Köhler was put in charge of it. The outbreak of World War I effectively marooned him, and Köhler remained on Tenerife until 1920, investigating the behavior of great apes.

In his research, Köhler believed he had found evidence of insightful problem solving in his nonhuman primates. In true gestalt fashion, he defined insight in perceptual terms—the sudden reorganization of the elements of the problem situation into an organized whole that solved the problem. His famous example was the two-stick problem, in which he placed a banana outside a cage, and the ape had to connect two hollow bamboo sticks (with different diameters) to reach it. After trying one stick at a time and failing, Sultan (Köhler's best-known ape) solved the problem when the sticks happened to line up as he was holding them. He pushed one into the end of the other and retrieved the banana (or so the story goes; it wasn't quite that simple; see Goodwin, 2005). On the basis of this and similar outcomes, Köhler argued that apes were capable of insight, and he criticized Thorndike's puzzle box experiments, writing that Thorndike never gave his cats the opportunity to perceive all the elements of the problem situation. Thorndike replied that Köhler's apes showed more trial and error than Köhler was willing to admit.

Gestalt psychology had an important influence on perception and cognition research. Modern introductory psychology textbooks always describe gestalt organizing principles in their perception chapter, and cognition texts acknowledge the debt to the gestaltists in their discussions of problem solving and their descriptions of the organizational attributes of memory.

Behaviorism

Behaviorism was a uniquely American school of thought that was very popular in the United States but had much less impact in other countries. As a school of thought, its founder was John B. Watson (1878–1958), a Johns Hopkins psychologist trained at the University of Chicago, a center of functionalism. Raised on a South Carolina farm, Watson always seemed to get along better with animals than with people, so it was natural for him to become interested in animal behavior. This interest eventually led him to believe that the definition of psychology as the science of mental life was a mistake; instead, he proposed, it should be a science of behavior. He made his ideas public in a 1913 talk at Columbia, *Psychology as the Behaviorist Views It*. Later that year, he published the talk, which has come to be called the "Behaviorist Manifesto" (Watson, 1913). In it, Watson argued that for psychology to become a genuine science, it needed phenomena that could be measured objectively. Introspection, central to the psychology of the day, was inherently subjective, however—introspector A could not observe introspector B's conscious experience, and vice versa. Behavior, however, could be observed objectively. Two observers could agree, for instance, that a rat had made three wrong turns in a maze. To be scientific, then, psychology ought to be the science of measurable behavior, according to Watson, and psychology's goal should be to understand the relation between environmental stimuli and the resulting responses. Watson also argued that behavior is mostly the result of the conditioning experiences of one's lifetime, and even claimed that if he had complete control of the lives of a dozen infants, he could shape them in any way he chose.

To demonstrate that environmental stimuli could condition behavior, Watson (with Rosalie Rayner) completed the famous "Little Albert" study in 1920, in which an 11-month-old boy was apparently conditioned to fear a white rat by pairing the rat with loud noise (Watson & Rayner, 1920). Although the study was seriously flawed methodologically, used only one subject, and could not be successfully replicated, it became the centerpiece of the behaviorist claim that people could only be understood by knowing their "conditioning history."

During the 1920s, several factors combined in such a way that, by the end of the decade, behaviorism had become the dominant school of thought in American academic psychology. First, the work of Russian physiologist Ivan Pavlov (1849–1936) became influential. Although Pavlov's research had been known about for years (he won a Nobel Prize in 1904 for his work on the physiology of the digestive system), his classical conditioning research became widely available for the first time in the English-speaking world when a large translation of his work appeared in 1927. Pavlov's insistence on a precise control of the environment and his model of conditioning that involved pairing conditioned and unconditioned stimuli blended perfectly with Watson's strong beliefs. A second factor was the continued promoting of behaviorism by Watson in the 1920s; this included the publication of several books for popular consumption, one on behaviorism in general and one on a behavioral approach to child rearing. The third development in the 1920s was the development, originally in physics, of operationism, which seemed to solve a problem with the issue of objectivity. The problem was that many of the phenomena of interest to behaviorists were not directly observable. Hunger, for example, was an important component of a behaviorist's maze learning study (nonhungry rats will not work hard to get to the goal box with food in it), but how does one know if a rat is hungry? The solution was to create what became known as *operational definitions*, definitions in terms of a set of operations that could be observed and measured. Hunger in a rat, for instance, could be operationally defined as 24 hours without food.

Pavlov's influence, Watson's continued proselytizing, and operationism opened the door for the era of behaviorism, which lasted roughly from 1930 until the late 1950s. During the period, there were plenty of nonbehaviorists continuing to research such topics as memory, attention, and perception, but the majority of academic psychologists found themselves completing research that tried to understand the essential nature of the conditioning process itself. Three psychologists in particular—Edward Tolman, Clark Hull, and B. F. Skinner—developed important theories of learning, and Skinner's work had impact even beyond the era of behaviorism. Skinner became one of the most prominent psychologists of the 20th century.

B. F. Skinner

By the time B. F. Skinner (1904–1990) delivered his last public address, a ringing defense of his ideas at the 1990 APA convention (given just eight days before he died of leukemia), he was the most recognizable name in American psychology. Skinner was one of a number of psychologists influenced by the events of the 1920s described earlier. After graduating from college as an English major in 1926, he spent a year at home trying to write. He failed, but in the process, he discovered popular accounts of behaviorism and read both Watson and Pavlov (Bjork, 1993). The experience led him to graduate studies at Harvard, where he earned a doctorate and began to develop his ideas. He referred to his system as the experimental analysis of behavior, and he focused his attention on a form of conditioning, distinct from Pavlovian or classical conditioning, that he called *operant conditioning*. Behaviors, he argued, are shaped by their immediate consequences and controlled by the environments in which these events occur. Positive consequences (e.g., positive reinforcement) strengthened future behavior, whereas negative consequences (e.g., punishment) weakened future behavior.

Skinner spent a lifetime elaborating the basic principles of operant conditioning, advocating for his system, and applying his principles to everyday life. His best-known laboratory research was on the so-called schedules of reinforcement, the various contingencies that occur when reinforcement does not follow every response. But he also became a public figure by writing a utopian novel, *Walden Two* (1948), that described a community based on operant principles, and by writing *Beyond Freedom and Dignity* (1971), which landed him on the cover of *Time* magazine. Skinner's hero was Sir Francis Bacon (1561–1626), and like Bacon, Skinner advocated an inductive approach to research (i.e., do lots of individual experiments, and then draw general principles from them), in the belief that an understanding of nature could be demonstrated by controlling nature. In the public eye, Skinner's ideas about controlling human behavior sounded vaguely anti-American. Thus, although he had enormous influence in psychology (operant conditioning will always occupy at least a third of the chapter on learning in general psychology texts) and his legacy is assured, his ideas were less influential in the broader social and cultural context.

Cognitive Psychology

Academic psychology began as the science of mental life, but research on human cognition suffered when behaviorism held sway during the 1930s and 1940s. During the 1950s, however, several factors combined to bring the study of cognitive processes (memory, attention, thinking, language, etc.) back into play. By the end of the 1960s, perhaps influenced by the broader historical context of that tumultuous decade, some writers were describing what they called a "cognitive revolution" in psychology. Although the change was more evolutionary than revolutionary, by the 1970s cognitive psychology had clearly supplanted behaviorism as psychology's reigning school of thought.

The reemergence of cognitive psychology was a product of forces both within and outside of psychology (Segal & Lachman, 1972). Within the discipline, a number of researchers began to question behaviorism's ability to explain such cognitive phenomena as human language. Considering the importance of language to the very concept of being human, this concern was a big problem for behaviorists. Outside of psychology, the rapid growth of computer science following World War II suggested a metaphor that would become central to the emerging interest in cognition. Specifically, just as computers received input from the environment, processed and transformed that input internally, and then produced some output, so did the brain receive input (sensation and perception), process the information (e.g., memory), and produce output (e.g., spoken language). Models of various cognitive processes started to look like computer flowcharts—memory chapters in general psychology books still show memory models with boxes for short- and long-term memory, and with arrows representing the flow of information—and research programs on cognitive processes proliferated. By the 1980s, the term *cognitive science* began to appear, reflecting the growth of an interdisciplinary approach that combined cognitive psychologists, computer scientists, philosophers of science, linguists, and anthropologists. Cognitive scientists have been interested in such topics as artificial intelligence. At the end of the 20th century, interest in cognition remained high.

THE PRACTICE SIDE: APPLIED PSYCHOLOGY

Despite E. B. Titchener's claim, early in the 20th century, that psychologists should concentrate their efforts on basic laboratory research, most American psychologists, raised in a country that valued practicality, were interested in how the new psychology could be applied to solve societal problems. Furthermore, although most traditional histories of psychology (e.g., Boring, 1929) have concentrated on the development of experimental psychology in an academic environment, historians now recognize that the history of applied psychology is of equal importance. Three prime examples of applied psychology in the 20th century were business or industrial psychology, the mental testing movement, and clinical psychology.

Industrial Psychology

It has been said that the business of America is business, and if that is the case, then it is not surprising that shortly after psychology began to think of itself as an independent discipline, applications of psychology to the world of business began to appear. In 1903, Walter Dill Scott (1869–1955), who was trained in traditional experimental psychology, launched a distinguished career in what was then called economic psychology by publishing *The Theory and Practice of Advertising*. In it, he argued that consumers were highly suggestible and that appeals to emotion could be effective as an advertising strategy (Benjamin & Baker, 2004). He followed this pioneering book in 1910 with *Human Efficiency in Business*, in which he made suggestions about how to improve productivity by applying psychological principles.

Hugo Münsterberg

Although Scott has a more legitimate claim, the German psychologist Hugo Münsterberg is sometimes considered the "founder" of industrial psychology. Another experimentalist by training, William James recruited Münsterberg in 1892 to run the Harvard laboratory. By the turn of the 20th century, however, Munsterberg had become more interested in applied psychology; in 1908, he published *On the Witness Stand*, often considered the pioneering text of forensic psychology. However, he was better known for his 1913 *Psychology and Industrial Efficiency*. This book was divided into three main sections, each of which has become a major area of modern industrial psychology. Section 1 concerned how the methods of psychological research could enhance employee selection procedures, section 2 showed how to improve productivity, and section 3 concerned advertising and marketing.

Scott and Münsterberg were just two of a number of groundbreaking industrial psychologists. Other important pioneers were Harry Hollingworth (1880–1956), Walter Van Dyke Bingham (1880–1952), and Lillian Gilbreth (1878–1972). Gilbreth, for example, managed to earn two doctorates, run a successful consulting business, and be a pioneer in the area of ergonomics, all while raising 12 children (her life was the basis for the 1949 book, *Cheaper by the Dozen*, written by two of her children).

The Mental Testing Movement

As mentioned earlier, Cattell created the term "mental test." His strategy, with its emphasis on measuring sensory capacity, failed and was abandoned. The modern approach to testing mental ability originated in France with the work of Alfred Binet (1857–1911). Binet and his colleague

Theodore Simon developed their test for the purpose of identifying which French schoolchildren would need extra academic attention. Rather than measuring sensory capacity, Binet and Simon created tests that assessed higher-level cognitive skills, especially those skills thought to produce success in school (e.g., given three words, place them in a meaningful sentence). The first test appeared in 1905; it was revised in 1911, shortly before Binet's death. The tests identified a student's "mental level." Five-year-olds who performed at a mental level of five were on target, whereas other five-year-olds who scored at a mental level of six were somewhat advanced in their learning. Binet recommended that students who scored two mental levels below their age should be in the special classes. It is worth noting that Binet did not think that intelligence was a fixed quantity, determined more by nature than by nurture. Rather, he was convinced that mental ability could be shaped by experience—given good teaching, the child two mental levels below their age would nonetheless be capable of catching up.

Henry Goddard

Henry Goddard (1866–1957) was a student of G. Stanley Hall, and he earned a doctorate from Clark in 1899. After teaching for a brief time, he became research director at the Vineland Training School for the Feeble-Minded in southern New Jersey. At Vineland, he initially tried to measure the mental abilities of the children housed there by using Cattell-like measures, but he soon found them to be useless. On a trip to Europe, he discovered Binet's work, and it is Goddard who was responsible for bringing the Binet ability tests to America. He used them at Vineland to classify students, and he became especially concerned about a particular kind of student—someone who appeared outwardly "normal" but was impaired mentally and, in Goddard's judgment, was responsible for many of society's problems (likely to become a criminal, a prostitute, etc.). These were primarily adolescents who scored at a mental level of between 8 and 12, and Goddard created the term *moron* to identify this group. In contrast with Binet, Goddard believed that mental ability was inherited (Mendelian genetics had just been rediscovered, and genetic explanations for a variety of behaviors were common). That made it important to be able to identify problem children and keep them out of trouble by maintaining them at places like Vineland, or so Goddard believed.

To bolster his claims about the genetics of intelligence, Goddard completed a famous 1912 genealogical study of a family that he named the Kallikaks (a combination of Greek words meaning "good" and "bad"). The study began when it was determined that a young girl at Vineland had numerous relatives in the region who seemed to divide into two groups—some were impoverished and likely to be involved in crime, and others were upstanding members of the community. Goddard and his assistants were able to trace the family line to a Revolutionary War soldier who had a casual affair with a feeble-minded woman in a bar, and later married into a fashionable family. Both unions produced children, which eventuated in the two different lines. Completely ignoring environmental factors such as poverty and privilege, arbitrarily labeling some individuals feeble-minded (e.g., the barmaid, apparently because she allowed herself to be seduced), and selectively using the evidence, Goddard believed he had uncovered clear evidence for the inheritability of intelligence. It is worth noting that later in his life, after much more was known of the complex interaction between heredity and environment, he expressed regret about his strong conclusions in the Kallikak study.

Lewis Terman

It is because of the influence of Lewis Terman (1877–1956) that the term *IQ* (intelligence quotient) is so widely used today. Whereas Goddard first brought the Binet tests to America and translated them, Terman undertook a systematic standardization of the test and produced the Stanford-Binet IQ test in 1916. It was later revised several times and is still in use. Although Terman did not originate the concept of IQ, his test popularized the idea that a single number could summarize someone's mental ability. The original formulation divided mental age (same concept as Binet's mental level) by chronological age and multiplied by 100.

Because Terman believed that intelligence was largely inherited, and that it required a certain amount of ability to be successful, he argued that American society ought to be a meritocracy—only the mentally capable should be in positions of authority. To identify such individuals, and to provide evidence to support his beliefs, Terman began a study of mentally "gifted" children in 1921. After identifying almost 1,500 high-IQ California children, he and his students tracked them for years. In fact, the study outlived Terman and became psychology's longest-running longitudinal study. Terman believed his study disproved the stereotype of the brilliant child who burned out in early adulthood and failed socially. Terman's participants (they called themselves "Termites") as a group remained highly successful throughout their lives, although the women in the group tended to be frustrated because of limited opportunities.

Robert Yerkes

The original passion of Robert Yerkes (1876–1956), one to which he returned near the end of his career, was the study of animal behavior (the Yerkes Primate Research Center in Atlanta is named for him). In between, he became a prominent promoter of intelligence testing, and because of World War I, he became a pioneer in the creation of group intelligence tests.

As the United States was beginning to enter the first world war, Yerkes was able to convince the military to authorize the creation of a group test that would help classify soldiers for different jobs, identify officers, etc. These tests eventually became known as Army Alpha and Beta. Alpha was given to soldiers who could read and write, whereas Beta was given to soldiers who were illiterate or failed Alpha. Most historians do not believe the tests had any effect on war readiness or the performance of the army, but Yerkes effectively promoted the project after the war as a great success, and the 1920s became a decade characterized by a virtual explosion in ability testing (including a test that was the forerunner of the modern SAT).

Clinical Psychology

Perhaps the most visible application of psychological principles concerns the care of those with mental and behavioral disorders. History's best-known therapist was Sigmund Freud, of course, although his training was in medicine, and he would not have called himself a clinical psychologist. Freud is well known to students for his three-part theory of personality (id, ego, superego), his stage theory of development (oral, anal, phallic), his work on the relation between anxiety and defense mechanisms (e.g., repression), and his overriding belief in the importance of the unconscious in determining our behaviors and ways of thinking.

In the United States, Lightner Witmer (1867–1956) of the University of Pennsylvania created the term *clinical psychology* and opened American psychology's first clinic near the turn of the 20th century. It is worth noting that Witmer's version of clinical psychology was closer to what would be called school psychology today. Indeed, the APA's division for school psychology gives an annual "Lightner Witmer" award to a promising school psychologist.

Clinical Psychology After World War II

The modern version of clinical psychology—the diagnosis and treatment of mental and behavioral disorders—emerged from the chaos of World War II. As the war progressed, it quickly became apparent that psychiatrists, up to then the ones primarily responsible for treating mental illness, could not handle the caseload. For instance, of the first 1.5 million soldiers given medical discharges, approximately 45 percent were for psychiatric reasons (Vendenbos, Cummings, & DeLeon, 1992). In this context, psychologists began to fill the void, and shortly after the war ended in 1945, the United States government started funding training programs for clinical psychologists. In 1948, a number of clinicians, led by David Shakow (1901–1981), met in Boulder, Colorado, and produced a training model that quickly became the gold standard for training clinical psychologists—the scientist-practitioner model (also called the Boulder model). It emphasized solid practitioner training while insisting on a research-based dissertation that led to a PhD.

With the Boulder model in place, clinical psychology grew rapidly in the 1950s, and new therapeutic strategies began to evolve. These new strategies were given impetus by a devastating critique of traditional Freudian psychotherapy. The British psychologist Hans Eysenck (1952) evaluated a number of studies of therapy effectiveness and discovered that traditional therapy did not seem to be any better than no therapy at all. Eysenck's methods came into question, but the damage was done. During the 1950s and 1960s, new therapies appeared. One came from the behaviorist perspective, and was called *behavior therapy* (e.g., if conditioning can explain the development of a phobia, then conditioning principles can be used to eliminate or unlearn the phobia). The second came from what was called psychology's Third Force, and a therapeutic approach associated with Carl Rogers.

Carl Rogers

Humanistic psychology, at least as it concerned clinical treatment, started as a revolt against psychoanalytic (Freudian) approaches to therapy and behavioral approaches. Although these two approaches are quite different in many ways, one shared feature was an emphasis on the past in determining and shaping a person's present. In contrast to these two "forces," humanistic ("Third Force") psychologists rejected the idea that repressed biological instincts or prior conditioning history determined a person's present and future. Instead, they argued that individuals were not tied to their past, that free will and responsibility were the attributes that most clearly characterized humans, and that virtually everyone could take control of their lives and realize their full potential as humans. Carl Rogers (1902–1987) was a leader in this movement.

Rogers earned his doctorate in psychology from Columbia in 1931, well before the Boulder model existed. His training was partly in child guidance, and his internship at a clinic with a psychoanalytic bias convinced him that Freud was not for him. His first academic position was at Ohio State University, and it was there that he developed a humanistic approach to therapy that he called client-centered therapy. It rejected the idea that delving into a client's past was essential for successful therapy, and developed the idea that the client could take responsibility for his or her improvement if the therapist did a good job of creating a safe, accepting, and empathic therapeutic atmosphere. Rogerian therapy, in turn, led to a number of other, similar approaches that were all under the umbrella of humanistic therapy. Even before the Boulder training model had been developed, Rogers proposed a training program for clinicians that weighed practice more heavily than research. This concept eventually produced a conference in Vail, Colorado, in 1973 and the development of the PsyD degree. If the Boulder PhD program can be considered a scientist-practitioner model, the Vail PsyD program is a practitioner-scientist model. Both models have APA approval.

AT THE CLOSE OF THE 20TH CENTURY

Over the last few decades of the 20th century, several trends developed in psychology, and they have carried over into the new century. There are at least four of them, and the following brief descriptions will conclude this chapter.

1. Neuroscience

Psychology has roots in 19th-century physiology, and psychologists have always been interested in the relationship between brain and behavior. In recent years, however, this interest has escalated dramatically, primarily due to increased technology (e.g., highly detailed brain scans).

2. Evolution

Darwinian thinking has also been central throughout psychology's history (e.g., the cornerstone of functionalist thinking), but recent years have seen the development of an entire subdiscipline called evolutionary psychology, which uses evolutionary principles to explain a number of human behaviors (e.g., sexuality). Textbooks have begun to appear, and courses on evolutionary psychology are multiplying rapidly.

3. Computers

The development of high-speed computing ability has had, and will continue to have, a dramatic effect on research in psychology. One effect has been a dramatic increase in the kinds of statistical analyses that have been done (e.g., more multivariate analyses); a second effect has been in the lab—computers are now essential in virtually all aspects of data collection.

4. Fragmentation

As the 20th century ended, psychology had become increasingly specialized and, as a consequence, increasingly fragmented—cognitive neuropsychologists, industrial psychologists, and school psychologists seemingly had little to discuss with one another. This increased specialization will inevitably continue, but it is important to remember that psychologists will always have one thing in common—their history.

REFERENCES AND FURTHER READINGS

Benjamin, L. T., Jr. (2000). The psychology laboratory at the turn of the century. *American Psychologist, 55,* 318–321.

Benjamin, L. T., Jr., & Baker, D. B. (2004). *From séance to science: A history of the profession of psychology in America.* Belmont, CA: Wadsworth.

Bjork, D. W. (1993). *B. F. Skinner: A life.* New York: Basic Books.

Boring, E. G. (1929). *A history of experimental psychology.* New York: Century.

Fernberger, S. W. (1931). The history of the psychological clinic. In R. A. Brotemarkle (Ed.), *Clinical psychology: Studies in honor of Lightner Witmer* (pp. 10–36). Philadelphia: University of Pennsylvania Press.

Goodwin, C. J. (2005). *A history of modern psychology* (2nd ed.). Hoboken, NJ: Wiley.

Hindeland, M. J. (1971). E. B. Titchener: A pioneer in perception. *Journal of the History of the Behavioral Sciences, 7,* 23–28.

Hothersall, D. (2004). *History of psychology* (4th ed.). New York: McGraw-Hill.

James, W. (1890). *Principles of psychology.* Boston: Henry Holt.

James, W. (1892). *Psychology (Briefer course).* Boston: Henry Holt.

Segal, E. M., & Lachman, R. (1972). Complex behavior or higher mental processes: Is there a paradigm shift? *American Psychologist, 27,* 46–55.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* New York: Macmillan.

Vandenbos, G. R., Cummings, N. A., & DeLeon, P. H. (1992). A century of psychotherapy: Economic and environmental influences. In D. K. Freedheim (Ed.), *History of psychotherapy: A century of change* (pp. 65–102). Washington, DC: American Psychological Association.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20,* 158–177.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology, 3,* 1–14.

Winston, A. S. (1990). Robert Sessions Woodworth and the "Columbia Bible": How the psychology experiment was redefined. *American Journal of Psychology, 103,* 391–401.

3

Psychology Into the 21st Century

Jay C. Brown
Texas Wesleyan University

As the old millennium closed, there was a shift from behavioral perspectives in psychology to cognitive perspectives (Goodwin, 2004)—that is, a shift from a strict reliance on empirical sources for knowledge to an acceptance of rationality as a source for knowledge. In the first decade of the new millennium, psychology is taking on another new look, or perhaps it is simply returning to whence it came. Psychology, and in fact all of society, is embracing spirituality, accepting faith as a once-again legitimate source of knowledge. The media is awash with words such as intuition and faith (Malcolm Gladwell's book *Blink*, a best seller, attests to the popularity of intuitive thinking), thus reminding us of this acceptance of non-empirical sources of knowledge. Advances in information technology, such as the Internet and the 24-hour news networks, have made psychology in all of its forms accessible to everyone. This increased globalization of psychology is mirrored in the most recent edition of the DSM (Diagnostic and Statistical Manual), which now acknowledges that mental health practitioners should be aware of cultural issues when diagnosing and treating mental illness.

At the close of the previous chapter, Goodwin presented a section titled "At the Close of the 20th Century" and discussed four trends in psychology that have carried over into the new century: neuroscience, evolution, computers, and fragmentation. This chapter includes a discussion of these four trends and concludes with a discussion of the future of psychology by examining the applications of psychology in the world today.

TRENDS IN PSYCHOLOGY

Neuroscience

The American Psychological Association (APA) declared the 1990s the "Decade of the Brain." An acceleration of neuroscience research marked the 1990s. For example, researchers discovered that well over half of the human genes are devoted to brain-related functions and those genes can influence a range of behaviors. Brain scans had long shown the structure of the brain. Crude neuroscience techniques of the past (such as ablation) had indicated which "areas" of the brain were responsible for which behaviors. The major recent advance, however, is the ability to scan the brain while the participant performs complex tasks. This technique reveals much more precisely which areas control which behaviors. At the turn of the century, a multidisciplinary initiative was formed (including the APA) creating the "Decade of Behavior" to bring social and behavioral science into the mainstream. This initiative was founded under the assumption that though an understanding of the brain is critical, it is ultimately human behaviors that are at the root of all accomplishments, problems, and solutions.

In addition to the recognition in the academic sphere, there has been an increased understanding of the importance of neuroscience research in the private sphere as well Articles dealing with brain functioning appear frequently both in magazines and the news; rarely does a week goes by without some new neuroscience development being reported by the cable news networks. Moreover, the Allen Brain Atlas (2006), created to be freely available to all, was funded totally with private donations by the philanthropist Paul Allen. This project was created because a private citizen realized the necessity for a comprehensive resource to which individual researchers could compare genetic patterns.

In an almost Orwellian development, it appears that MRI scans of the brain can predict at least some psychological characteristics. The brains of highly intelligent children develop in a unique pattern. Though this pattern is quite complex, the pattern for highly intelligent children is definitely different from that for children with average intelligence (Haier et al., 1995; Shaw et al., 2006). Because of this different pattern, it should be (at least theoretically) feasible to "predict" the intelligence of an unborn child. Brain differences in personality, temperament, abnormal behaviors, and other cognitive functions have been reported for years, and more are sure to be found shortly. Obviously, serious ethical discussions need to be made in this area, particularly in the potential for prediction. If a person's potential is known beforehand, why waste precious resources on people who cannot benefit from them? Similarly, if the potential for violence in an individual can be assessed through brain scanning, should society prevent such "potential" violence from occurring? Not surprisingly, such a scenario was proposed in the recent hit movie *Minority Report*, in which crimes could be predicted before they occurred and the criminals were arrested for crimes they had not yet committed. Neuroscience may be the route that will allow psychology to finally fully advance beyond the realm of social science, where predictions are probabilistic, and into the realm of the natural sciences, where prediction is more of an absolute. Rather than merely predicting the behavior of the "average" individual, neuroscience may someday allow us to predict the behavior of a specific individual.

Evolution

Since its inception, the theory of evolution has been woven into the history of psychology. The Structural School of Psychology, which had asked about the "what," "why," and "how" of mental life, was supplanted by the Functionalists, who incorporated evolutionary thinking to ask about the "what for" of mental life (Hergenhahn, 2005). In the 1960s, the field of psychobiology (currently called evolutionary psychology), in an effort to explain altruistic behavior, changed the basic driving force behind evolutionary thinking from the concept of "survival of the fittest" (implying that characteristics that allow an individual to survive and reproduce would be passed on) and replaced it with the propagation of an individual's genetic material (which could be enhanced by ensuring that ones with similar genetics survive and reproduce).

A recent cover story in *Time* magazine asks, "What makes us different?" and refers to those characteristics that separate us from other animals, namely chimpanzees (Lemonick & Dorfman, 2006). The human genome project was launched in 1986, and a preliminary count reveals only about 30,000 genes in the entire human genome. Of these 30,000 genes, we appear to share about 99 percent with chimpanzees (*Pan troglodytes*), our nearest ancestor. Similarly, as of this writing, work is in progress to reveal the genome of Neanderthal man (*Homo sapien Neandertalis*). It is expected that we will share even more genes than we do with the chimpanzee. Full understanding of the human genome should reveal important insights into behavioral, mental, and physical illnesses. Early applications include the ability to screen for predispositions for certain disorders, such as breast cancer. Such "predictions" have opened other ethical debates that are being played out in medicine today. For example, if a woman's genetic testing reveals a predisposition for breast cancer, should she have a radical mastectomy just to be safe? Taken further, if genetic testing reveals predispositions for any number of illnesses (or cognitive impairments) in a fetus, should the fetus be given a chance?

Computers

From the information-processing approach to memory of the 1960s, to the connectionist models of language in the 1980s, to the field of artificial intelligence and neural networks today, computers have been influencing the way we think about and study the mind and its processes. However, there is a consensus among researchers that current computer technology, which is very, very fast, is limited in its ability to truly imitate a human brain due to the current limit of serial processing. Until a true parallel processing system can be devised, the field of artificial intelligence will forever remain artificial. In the words of Donald Hoffman (1998), "You can buy a chess machine that beats a master but can't yet buy a vision machine that beats a toddler's vision" (p. xiii).

Computers have also been shaping psychology via their roles in research. For example, computerized presentation of test items allows for far more standardization and better control in research than was ever possible in the past. Additionally, statistical software programs have allowed researchers to pursue complex statistical relationships that might otherwise be lost.

Fragmentation

Long gone are the days of the "gentleman scientist" of the 19th century. The trend of the past century, in which psychology has become more and more fragmented (specialized), looks only to continue into the future. The APA currently has 54 divisions, with more inevitably to follow, each specialized in its own small slice of psychology. We find that researchers in one area of psychology might belong to a very specific society such as the Society for Research in Child Development (SRCD), which uses very specialized terminology that is unrecognizable to someone researching in a different area of psychology. Researchers in one area can hardly keep up with advances in their particular slice of the research world. Human beings are incredibly complicated, and it feels hardly possible to understand one area of psychology in isolation from any of the others. However, evidence of an increased amount of interdisciplinary cooperation (particularly between psychology, biology, and computer science) is encouraging. Academic psychologists are often no longer fully recognizable as psychologists; the bookshelves of a psychologist might be filled with books written by anthropologists, linguists, economists, paleontologists, and so forth. It has been suggested that B. F. Skinner was perhaps the last completely recognizable figure in the field of psychology.

The Future

Psychology, like all sciences, must ultimately produce useful knowledge, knowledge that can benefit all of society. Psychology's usefulness and acceptance by the general public has been rising. We have seen a trend of an increasing merger of psychological research and ideas into the mainstream of 21st-century life. Psychology is being applied to almost all aspects of our daily lives with little understanding by the general public that it even is psychology. One particularly clever integration of psychological research into the mainstream would have made B. F. Skinner very proud—the Fisher-Price *Throne of Their Own*, a toilet-training chair that uses music to reinforce successful elimination. Similarly, *Newsweek* recently published an article titled "How to Read a Face" discussing an emerging new field called social neuroscience (Underwood, 2006). Perhaps the most interesting part of this article is that the word "psychology" is not mentioned even once! Reminiscent of George Miller's impassioned plea in his APA presidential address that we give psychology away, it seems that the very success of psychology could lead to its demise as a separate discipline and its fusion with mainstream thinking. In the following sections, we briefly highlight a few of those areas in which psychology (and the research tools it has created) has gained considerable influence.

PSYCHOLOGY AND OPINIONS

Consumer Behavior

The foremost name in the history of psychology related to consumer behavior is John Watson (Goodwin, 2004). Watson was credited with implementing classical conditioning principles at the J. Walter Thompson Advertising Agency that he originally developed with Rosalie Rayner while studying Baby Albert. Today we are constantly bombarded with new products (CS) being paired with people or places (US) that have positive feelings (UR) associated with them. Through constant bombardment (acquisition), the companies hope that the presence of the new product (CS) will eventually come to elicit positive feelings (CR).

When it comes to the development of new products, no successful company would even attempt to market a new product without testing it. The research techniques of polling and focus groups come directly out of psychological research. The availability of mass audiences via the Internet and e-mail makes this type of work faster and cheaper (though not necessarily better).

Politics

Political strategists have taken the psychological tools of polling and focus groups and created a new art form. Sometimes it appears as though the image created by a politician is more important than what is actually being said or done (especially in the middle of hot political campaigns). Everything a candidate plans is shown first to a focus group to find how it would affect people's opinion of the candidate. Newer techniques involving micro-polling allow political strategists to predict with much more precision which candidate each household is likely to vote for. This knowledge is then used for concentrated "get out the vote" efforts.

PSYCHOLOGY AND THE LEGAL SYSTEM

The presence of psychological principles in legal areas is hardly news. The insanity defense was one of the first applied uses of psychological principles in any domain. However, newer developments, such as the rash of "recovered memory" trials in the 1990s, bring to light the basic notion of the legal system—namely, that the words that come out of a witness's mouth (for either the defense or the prosecution) are supposed to be a valid reflection of reality.

In a strange twist highlighted by the recent movie *Runaway Jury*, we are reminded that the truth of a situation is of little importance when considering the outcome of a trial. In the movie, psychologists were employed to ensure that the jury members who were selected from the pool of potential jurors were sympathetic to the prosecution or defense, as the situation dictated. Situations such as these remind us that as psychologists we must maintain high ethical standards that go far beyond patient confidentiality.

PSYCHOLOGY FOR PREDICTION AND ASSESSMENT IN EDUCATION AND INDUSTRY

Beginning with the first large-scale use of psychological tests by Yerkes and his colleagues during World War I to test draftees' intellectual functioning (the modern *ASVAB*

[Armed Services Vocational Aptitude Battery] is a testament to the ongoing efforts in this vein), employers and academic institutions have been employing psychological tests at an ever greater pace. Despite protests concerning the "fairness" of aptitude tests (such as the SAT, ACT, or GRE in higher education), they continue to remain useful and valid predictors of job and academic success. Other predictive uses of psychological testing, including the uses of the *Myers-Briggs Type Indicator (MBTI)* and the *Minnesota Multiphasic Personality Inventory (MMPI)*, are also plagued with questions of validity and fairness (as well as proper administration). It is likely that psychological tests will have their largest future impact in the area of assessment. More and more, employers and educational institutions want "proof" that both new and accepted techniques are working. Proper implementation of the No Child Left Behind Act of 2001, the largest change in education in a generation, requires assessment techniques in order to determine which schools reach their goals. Such developments guarantee the future of psychological assessment techniques.

THERAPY

In therapy, changes have been brewing for several generations. Traditional psychoanalytic therapists are now a minority, their ranks being overshadowed by cognitive therapists. The cognitive approach to therapy empowers patients by putting them in charge of their own recovery, though of course it can only be applied to a certain range of disorders.

Similarly, there have been shifts in who provides therapy. The cost of medicine has skyrocketed in recent years, and the rise of health management organizations (HMOs) brought with it a push for "cheaper" mental health care. Therapy, once dominated by psychiatrists, is now practiced by psychologists and therapists as well as by physicians and members of the clergy. Also, a shift from individual therapy to group therapy is evident. The shift in mental health care to "cheaper" methods and providers has allowed access to mental health care by a larger portion of the population than had ever been possible in the past. Along with these shifts come further questions of cost efficiency driven by the HMOs, namely prescription privileges. Traditionally only psychiatrists have been allowed to prescribe medications, but there is now a shift toward allowing these privileges to be provided to licensed clinical psychologists as well (currently Louisiana and New Mexico give clinical psychologists prescription privileges, but several other states have laws in the making).

SUMMARY

In the fast-paced world of the late 20th and early 21st centuries, in which economic decisions are made based on the "bottom line," it seems inevitable (i.e., the *zeitgeist* is right) that Francis Bacon's (1620/1994) notions at the beginning of the scientific revolution should come full circle. "Human knowledge and human power meet in one; for where the cause is not known the effect cannot be produced" (p. 43). All fields of science, in order to continue to exist, must provide useful information. As seen in the mission statement of the APA (2006), "The objects of the American Psychological Association shall be to advance psychology as a science and profession and as a means of promoting health, education, and human welfare." Perhaps psychology is finally starting to grow up enough to prove its usefulness.

REFERENCES AND FURTHER READINGS

Allen Brain Atlas. (2006). Retrieved December 12, 2006, from http://www.brainatlas.org/aba/

American Psychological Association. (2006). *About APA.* Retrieved December 12, 2006, from http://www.apa.org/about/

Bacon, F. (1994). *Novum organum* (P. Urbach & J. Gibson, Eds. & Trans.). La Salle, IL: Open Court. (Original work published 1620)

Goodwin, C. J. (2004). *A history of modern psychology*. Hoboken, NJ: Wiley.

Haier, R. J., Chueh, D., Touchette, P., Lott, I., Buchsbaum, M., Macmillan, D., et al. (1995). Brain size and cerebral glucose metabolic rate in nonspecific mental retardation and Down syndrome. *Intelligence, 20*, 191–210.

Hergenhahn, B. R. (2005). *An introduction to the history of psychology* (5th ed.). Belmont, CA: Thomson-Wadsworth.

Hoffman, D. D. (1998). *Visual Intelligence*. New York: W. W. Norton.

Lemonick, M. D., & Dorfman, A. (2006, October 9). What makes us different? *Time*, *168*(15), 44–53.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., et al. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, *440*(7084), 676–679.

Underwood, A. (2006, October 23). How to read a face. *Newsweek*, *148*(17), 65.

4

Women and Minorities in Psychology

Alexandra Rutherford
York University

Wade Pickren
Ryerson University

> Feminist scholarship has repeatedly demonstrated that how and what we come to know depends on *who* we are. —*Morawski*, 1990, p. 175

In July 1892 well-known Clark University psychologist G. Stanley Hall met with a small group of his peers and founded the American Psychological Association (APA). At their first official meeting the following December, 31 additional members were voted in; all were white, and all were male (see Fernberger, 1932). However, as psychology grew throughout the first half of the 20th century, the proportion of women in the field increased. In 1946, psychologist Alice Bryan and her colleague Edwin Boring conducted a survey of American psychology and found that of the 2,672 doctoral-level psychologists who responded to their survey, 24 percent were women (Bryan & Boring, 1946). At this point in history, very little attention was paid to the representation of non-white psychologists in the field.

The proportion of women remained relatively stable until the late 1960s. The greatest growth in the numbers of women in psychology began in the early 1970s, largely due to the impact of the second wave of the women's movement, and has continued steadily since that time. Whereas in 1960, 17.5 percent of all doctoral degrees in psychology in the United States were awarded to women, by the year 2000 the proportion of women receiving doctorates in the field had risen to 66.6 percent (Women's Programs Office, 2003). Psychology, especially its applied branches, is quickly becoming a female-dominated profession.

The 1960s also saw important cultural and political shifts that affected the number and representation of minority psychologists. The civil rights movement and the development of black nationalism provided the cultural and political foundations for the institutional and theoretical challenges of black psychologists. Their activism paved the way for other minority groups such as Latino/Latina psychologists and Asian American psychologists to demand greater receptivity to and support for their concerns and agendas within a largely white, Eurocentric psychological establishment. Although growth in numbers has been slow relative to the influx of women psychologists, ethnic minority men and women psychologists steadily continue to challenge and change the institutional, theoretical, and practical bases of the field.

CHAPTER OVERVIEW

In the first half of this chapter, we focus on the history of women and feminism in American psychology. First, we briefly survey selected works by early women psychologists intent on using the tools of the new science of psychology to challenge prevailing stereotypes of and biases towards women. Then we describe the status of women in the profession during the middle of the century, with a focus on activities around World War II. We conclude by examining the profound effects of second-wave feminism on both psychological theory and practice, and the institutional structure of psychology.

In the second half of the chapter we turn to the history of ethnic minority psychologists. The lives and contributions of select early African American pioneers are highlighted, but we note that American psychology attracted few minority psychologists and remained oblivious to their concerns and critiques until fairly recently. It took strong advocacy from individuals and groups, starting in the late 1960s, to make psychology more receptive to ethnic minority issues. The events in American psychology that led to these changes and the impact of ethnic minority psychologists on the processes and products of psychology are then discussed.

It should be noted that this chapter is limited to developments in the American context. Students interested in the history and status of psychologies developed in other countries and efforts by psychologists throughout the world to develop theory and praxis that are relevant to their local contexts can consult a growing body of literature, including Brock (2006), Kim and Berry (1993), and Paranjpe (2002). We again emphasize that *who psychologists are* impacts *what we know* in psychology. Accordingly, as indigenous psychologies develop around the world, and as American psychologists become more diverse, the potential for generating psychological knowledge that relates to and illuminates the diversity of the *entire* human experience becomes realizable.

WOMEN IN PSYCHOLOGY

First-Wave "Feminist" Psychologists

Women have been participating in psychology since its inception as a formal scientific discipline in Europe and North America in the late 1800s. Although women were often excluded from men's elite professional circles such as the founding of the APA and E. B. Titchener's Society of Experimentalists, they nonetheless made important intellectual and institutional contributions despite their small numbers and the considerable obstacles they faced in a male-dominated profession. Although it is not possible here to discuss all of the first-generation American women psychologists, Scarborough and Furumoto (1987) have identified 25 women who were members of APA or who listed their field as psychology in *American Men of Science* by 1906. Two of these women, Mary Whiton Calkins and Margaret Floy Washburn, served as presidents of the APA in 1905 and 1921, respectively. Another 50 years would pass before another woman, Anne Anastasi (1908–2001), would be elected president.

In this section, we focus on three early psychologists who explicitly challenged social expectations and stereotypes about women. Helen Bradford Thompson Woolley and Leta Stetter Hollingworth used the tools of psychological science to examine and refute commonly held beliefs about women's inferiority. Christine Ladd-Franklin campaigned vigorously for egalitarianism in professional conduct and exchange among psychologists, and equal access to education and honors for women.

Helen Bradford Thompson Woolley

Woolley (1874–1947) received her PhD in 1900 from the University of Chicago. For her dissertation research she conducted one of the first comprehensive empirical examinations of sex differences in intellectual, motor, sensory, and affective abilities (Thompson, 1903). In this work, she repeatedly found more similarities than differences between her male and female participants. When differences did emerge she interpreted them in light of the differential experiences and training of men and women, especially during early development. She was critical of the tendency to minimize environmental influences in favor of instincts and biology. In her conclusion to the work, she noted, "The psychological differences of sex seem to be largely due, not to difference of average capacity, nor to difference in type of mental activity, but to differences in the social influences brought to bear on the developing individual from early infancy to adult years" (Thompson, 1903, p. 182).

After earning her doctorate, Woolley went on to occupy a number of positions including director of the Bureau for the Investigation of Working Children in Cincinnati, where she formulated research and policy on child welfare reform (see Milar, 1999). When her husband moved to Detroit, Woolley gained a position there at the Merrill-Palmer School, where she initiated one of the first nursery schools for the study of child development. At Merrill-Palmer she continued her work on mental abilities by examining the mental abilities of young children. Finally, with a strong record of research and leadership in child development established, Woolley was offered an appointment as director of the new Institute of Child Welfare Research and professor of education at Teachers College, Columbia University. Unfortunately, this appointment and her move to New York marked the beginning of a number of emotional and physical traumas, including isolation from her daughters and friends, permanent separation from her husband, and a hysterectomy. In combination, these events precipitated her premature retirement in 1930. She lived with one of her daughters for the remaining 17 years of her life.

Leta Stetter Hollingworth

Leta Stetter Hollingworth (1886–1939) also used her scientific training to challenge widespread beliefs and stereotypes about women (see Shields, 1975). In her dissertation research under the supervision of Edward Thorndike at Columbia University, she undertook an empirical investigation of functional periodicity, the commonly held belief that women became psychologically impaired during menstruation. Hollingworth's study disconfirmed this belief, showing that women's perceptual and motor skills did not vary as a function of their monthly cycles.

Later in her career she addressed another widely held theory about the sexes: that men exhibited greater variability than women across all psychological and physical traits. According to the variability hypothesis, the males of the species were presumed to drive evolutionary progress. Only men were deemed capable of the most impressive intellectual, social, and political achievements. In short, it held that because women as a class exhibited less variability and range in their abilities, they were doomed to mediocrity. Unsatisfied with the evidence for this view, Hollingworth conducted several studies to determine if the variability hypothesis would be supported. Needless to say, she did not find empirical support for greater male variability and criticized adherence to the belief on both empirical and interpretive grounds. She maintained that the true potential of women could only be known when women received complete social acceptance of their right to choose career, motherhood, or both.

Christine Ladd-Franklin

Christine Ladd-Franklin (1847–1930) was an eminent first-generation woman scientist and an outspoken advocate for the equal treatment and acknowledgment of women in professional spheres. Ladd-Franklin attended newly established Vassar College for women as an undergraduate, where she majored in science and mathematics. She then taught these subjects in secondary schools for over a decade while continuing to study mathematics and publish papers in this field. When Johns Hopkins University opened in Baltimore, Ladd-Franklin eagerly applied to do advanced work in mathematics. Johns Hopkins was the first university in the United States to focus on research and graduate training, but it would not allow women students. However, a mathematics professor who was familiar with her papers permitted her to attend his lectures in the 1878 to 1879 academic year. As her reputation grew, Ladd-Franklin was given access to further courses and completed all the requirements for a PhD in mathematics and logic by 1882. It was not until 1926, however, when she was almost 80 years old, that Johns Hopkins would award her the degree.

Soon after the completion of her doctoral work, Ladd-Franklin's interests turned to psychology. When her husband traveled to Germany for a sabbatical, she arranged to spend six months studying color vision in G. E. Müller's laboratory in Gottingen and the next six months in Hermann von Helmholtz's laboratory with Arthur König, a physicist interested in color vision. As a result of this work, she produced her own theory of color vision and quickly established herself as a world authority on the topic (see Furumoto, 1992).

Throughout the rest of her career, in addition to promoting her theory, giving papers, and delivering lectures, she continued to advocate for women's equal treatment in the professions. She criticized the American Academy of Arts and Letters for not admitting women. She also took up a cause much closer to home: E. B. Titchener's exclusion of women from his invitation-only Society of Experimentalists. The Society, established by Titchener in 1904, was formed to promote discussion of experimental work among leaders in the field, junior faculty, and promising graduate students. Titchener's adamance that the group exclude women on the grounds that men would not feel free to engage in freewheeling critique—and smoke—in the presence of the "weaker sex" was met with incredulity by Ladd-Franklin. In a series of letters to Titchener, Ladd-Franklin expressed her view that this practice was both immoral and unscientific. Her attack was largely unsuccessful; it was not until after Titchener's death in 1927 that women were invited to join the group.

Women in Psychology at Mid-Century

Woolley, Hollingworth, and Ladd-Franklin lived and worked during a period when women were making important strides in American society. The suffrage movement had secured women the right to vote in the United States by 1919. Thus, their work in psychology was buttressed by what is now called "first-wave feminism." The middle of the 20th century, especially the period leading up to and just after World War II, saw much less attention paid to the rights of women and the status of women in society. Accordingly, this was a period in psychology when women continued to face many of the same obstacles as their first-generation counterparts, but without the support of a cultural milieu that was sympathetic to these challenges. Here we discuss the status of women in psychology in this period and present brief accounts of two attempts to draw attention to the problem of women's exclusion: the formation of the National Council of Women Psychologists, and the collaboration of Alice Bryan and Edwin Boring on "the woman problem."

National Council of Women Psychologists

In times of war and other national and international emergencies, most professional groups organize to offer their services. Psychology has been no exception. In World War I, for example, psychologists helped the United States Army make personnel decisions by developing and administering group intelligence tests designed to sort thousands of recruits into appropriate positions within the military. With the outbreak of World War II, psychologists once again organized to offer their services. In 1940, the Emergency Committee in Psychology (ECP) was formed to oversee the mobilization of psychologists in the war effort. A problem soon arose: the all-male ECP chose to mobilize only male psychologists.

When it became apparent that the expertise of women psychologists was not being called upon, about 30 women psychologists, all members of the American Association of Applied Psychology (AAAP), confronted the AAAP representative to the ECP, Robert Brotemarkle, with their concerns. Although sympathetic, Brotemarkle admonished the group to be "good girls"—to be patient and to wait quietly until plans could be made that would include them (Schwesinger, 1943). When, almost two years later, nothing had been done to include women, a group of about 50 New York women psychologists began meeting to discuss how they could use their professional skills in the national emergency.

In November, a subgroup of these women met in the Manhattan apartment of psychologist Alice Bryan to draw up a charter for a national organization of women psychologists. On December 8, 1941, just a day after news of the bombing of Pearl Harbor, the National Council of Women Psychologists (NCWP) was formed. Florence Goodenough, a respected psychological scientist, was selected as president. Although not particularly sympathetic to the plight of women as a group, Goodenough did support the goal of applying psychological expertise to needed areas. By the middle of 1942, 234 doctoral-level women psychologists had joined the NCWP (Capshew & Laszlo, 1986).

Although the impetus for the formation of the NCWP was the exclusion of women from the war effort, women psychologists were reluctant to make the NCWP solely a clearinghouse for charges of sex discrimination. As Alice Bryan (1986) remarked, in the devastating aftermath of Pearl Harbor, "Winning the war had to be given first priority" (p. 184). In addition, male psychologists and leaders in the ECP used subtle strategies to undermine the women's feminist resolve (Capshew & Laszlo, 1986). Many male psychologists denied that sex discrimination existed in psychology and suggested that in drawing attention to gender issues at a time of national emergency women were being self-indulgent. Others suggested that professional success in science was completely meritocratic; therefore, women had no right to demand special consideration on the basis of sex. Nonetheless, the formation of the NCWP marked the first time women had come together with the explicit aim of professional advancement.

"The Woman Problem"

In 1951, prominent Harvard psychologist Edwin Boring published an article in the *American Psychologist* titled "The Woman Problem." In this short, expository article, Boring noted that in terms of their positions in APA, "professional women acquire less prestige than professional men 'in proportion to their numbers'" (p. 679). He then suggested that two of the primary reasons for women's lower prestige were (a) a natural predisposition in women for "particularistic" tasks over the work of generalization and theory building that was the true calling of the scientist, and (b) job concentration difficulties; that is, women chose more often to fulfill family obligations over work obligations. He concluded that because programmatic scientific work and fanaticism were generally rewarded, it was not surprising that women would experience conflict between professional success and family orientation. Finally, he tackled the question of whether a woman could become such a fanatic and still remain marriageable. He concluded that indeed she could, but that she must be "abnormally bright to combine charm with sophistication" (p. 681).

Boring's single-authored article was actually the culmination of a somewhat complex collaboration with his colleague, Columbia University psychologist Alice Bryan, on the so-called woman problem. Bryan met Boring in 1942, when she was the only woman member appointed to a committee charged with reorganizing the APA. Boring was provoked by Bryan's repeated assertions that women did not hold representation in APA offices proportionate to their numbers in psychology and suggested that they collaborate on a study of the problem. Their study resulted in three articles published in the *Psychological Bulletin* and the *American Psychologist* between 1944 and 1947 (Bryan & Boring, 1944, 1946, 1947).

In his autobiography, Boring (1961) described his motive behind the collaboration. He hoped that Bryan, with her feminist convictions, and he, with his conviction that for both biological and cultural reasons women "determined most of the conditions about which she complained" (p. 72), might moderate each other's ideologies and get closer to the facts. Ultimately, in an effort to work together amicably the pair sidestepped their ideological differences and presented results that vindicated Bryan's suspicions that women were underrepresented, but avoided providing any interpretation of these findings. Boring's 1951 article, however, clearly conveyed his own interpretation: while acknowledging that disparities did exist, Boring was unwilling to attribute the disparities to discrimination.

As many feminist scholars have noted, despite the work of Alice Bryan and the NCWP, the postwar period of the 1950s did not provide a hospitable environment in which to sustain feminist activism. When the war ended, the NCWP was renamed the International Council of Women Psychologists. The reorganized group adopted a distinctly apolitical mission statement, began to admit more men, and was renamed the International Council of Psychologists (see Walsh, 1985). Unfortunately, the economic prosperity of the postwar period did not necessarily benefit women psychologists. Men quickly filled both the academic positions they had previously dominated and the newly prestigious clinical positions that had been at least partially situated in the "women's sphere" before the war. It was not until the second wave of the feminist movement in the late 1960s that organized psychology would again have to examine its professional treatment of women and confront issues of sex discrimination.

Second-Wave Feminism and Psychology

In 1963, Betty Friedan (1921–2006) published *The Feminine Mystique*, ushering in a period of second-wave feminism in the United States. For the next decade, feminist psychologists waged their own battle with their chosen discipline, demanding that sexist practices and androcentric theories be acknowledged and reformed. One of these psychologists was Naomi Weisstein, a Harvard-trained cognitive scientist and founder of the Chicago Women's Liberation Union. In the fall of 1968, she delivered a paper that was destined to become one of the founding documents of feminist psychology. Originally published by the New England Free Press in 1968 as "Kinder, Kirche, Küche as Scientific Law: Psychology Constructs the Female," a revised and expanded version was published in 1971 and has since been reprinted dozens of times in a wide range of publications (Weisstein, 1971).

In this article, Weisstein argued that psychology had nothing to say about what women were really like because, essentially, psychology did not know. This failure was due to psychologists' focus on inner traits and consequent ignorance of social context, as well as their failure to consider evidence. In this section, we explore the institutional changes that occurred in psychology as a result of second-wave feminism and the activism of feminist psychologists. We then briefly describe feminist psychology's internal struggle to embrace diversity and be relevant to the lives of all women.

Feminists Challenge the APA

In 1969, emboldened by the cresting of the women's movement and the pioneering efforts of psychologists such as Weisstein, feminist psychologists met at the annual convention of the APA where "regular symposia became angry discussions focused on sexist practices at the convention" (Tiefer, 1991, p. 636). These sexist practices included job advertisements indicating "men only," lack of child care at the convention, and overt sexual harassment. The result of these angry discussions was the formation of the Association for Women in Psychology (AWP) in 1969.

Members of the newly established AWP rallied again at the 1970 convention, presenting their concerns to APA president Kenneth B. Clark at an explosive Town Hall Meeting. Two pioneer feminist psychologists, Phyllis Chesler and Nancy Henley, prepared a statement on APA's obligations to women and demanded one million dollars in reparation for the damage psychology had perpetrated against women's minds and bodies. Dorothy Riddle, a spokesperson for the AWP, accused the APA of using stalling tactics with the AWP's demands rather than addressing its own sexism.

In response to these challenges, APA established a Task Force on the Status of Women, chaired by Helen Astin. The task force undertook a two-year study and published a detailed report of its findings and recommendations in 1973. One of these findings was that psychological research on and knowledge about women was deficient. Accordingly, the task force recommended that a division devoted to the psychology of women be established to promote research in this area. An Ad Hoc Committee on the Status of Women was formed to follow through on the recommendations of the task force, and in 1973, Division 35, Psychology of Women, was formed. Elizabeth Douvan was the first president of the division (see Mednick & Urbanski, 1991; Russo & Dumont, 1997). In 1976, the first issue of the new journal, *Psychology of Women Quarterly*, appeared. By 1995, Division 35 had grown to become the fourth largest in the APA.

Feminism, Diversity, and Inclusiveness

The formation of AWP and Division 35 was largely driven by a fairly homogeneous group of white, middle-class feminists who resonated with liberal feminism and had already established an institutional presence in psychology. Women of color have not had a strong institutional base in psychology and have not had a ready ear for their concerns. Early on, however, feminist psychologists recognized the need to attract the interests and energies of feminists of color and to make the psychology of women a field that would reflect the richness of this diversity (see Comas-Diaz, 1991). Moreover, it was recognized that parallels between racism and sexism could not be overlooked.

In 1976, the president of Division 35, Martha Mednick, asked Saundra Rice Murray to organize a task force on black women's concerns. The task force compiled a bibliography of research on black women, organized convention programs on the concerns of women of color, and worked to increase the representation of black women in APA governance. In 1978, the task force became the Committee on Black Women's Concerns, with Pamela Trotman Reid as its first chair. In 1985, a bylaw change converted the Committee to a permanent Section on the Psychology of Black Women, and Reid became the division's first black woman president in 1991. Since then, sections on the Concerns of Hispanic/Latina Women and Lesbian and Bisexual Women's Concerns have been formed. Women of color are slowly reaching the highest levels of APA governance, but despite this, as of 2006, no woman of color had been elected APA president.

Obviously, the history of women in psychology and the development of feminist psychology is a history-in-progress. Feminists have encountered both failures and successes in their attempts to mesh feminism and psychology. Ongoing challenges include responding to and incorporating more radical forms of feminist critique such as third-world and racialized feminism, and incorporating a range of feminist methodologies. As *who psychologists are* changes, so too must the methods and theories of the field.

MINORITIES IN PSYCHOLOGY

In this section, we focus on the history of the involvement of racial and ethnic minorities in the field of psychology in the United States. Histories of minorities in other countries and cultures, though few in number, would provide a useful comparison with our work here (e.g., see the chapters in Bond, 1997; Louw, 2002; Richards, 1997). As with all histories, this is a transitional document. Given the changing demographics of the United States, which indicate that Americans of European descent will be in the numerical minority by the year 2050, our writing here is very provisional. A history of the field written one hundred years from now may well have a very different focus from the one discussed here.

When G. Stanley Hall founded the American Psychological Association in 1892, those included in the initial membership represented a range of scientific and scholarly fields: psychology, philosophy, physiology, medicine, and so forth. Yet, not one of the original members, or any of the members for many years to come, was from a racial or ethnic minority. This is not surprising, as there were very few scientists or scholars at this time who were minorities. The reasons for this scarcity are both complex in detail yet, in some ways, simple. The United States was a society built on a foundation of white elitism.

Educational systems favored whites of Anglo-Saxon or northern European descent. However, as schooling became compulsory, opportunities for education for most of the population increased. For children of minorities, those opportunities were different in kind and quality, with fewer resources devoted to the education of African American, Native American, Latino/a, or Asian American children. Nevertheless, with the dawn of the 20th century, the issue of race and ethnicity loomed larger. As the noted black scholar and sociologist W. E. B. DuBois so famously wrote in 1903, “the problem of the Twentieth Century is the problem of the color-line” (p. 1). The history of that color line in psychology centers on the determination of people of color to find a place in the new field and to shape a body of knowledge and practice that is true to their values and heritage.

The authoritative history of the early years of this struggle has been written by the African American psychologist-historian, Robert V. Guthrie (1998). As Guthrie points out, blacks simply did not appear in standard histories of psychology such as those by Boring or Murphy. Within the field, racial minorities did appear as subjects, typically in research reports about racial differences. Such research typically produced results that were interpreted as indicating the inferiority of African Americans or Chicanos (Americans of Mexican descent) or some other minority (Richards, 1997).

Despite the lack of inclusion of racial and ethnic minorities in the histories of the field, there were, of course, individuals who were ethnic minorities who were involved in applied psychology. Few opportunities existed for higher education for minorities at the beginning of the 20th century, and fewer still for work beyond the bachelor’s degree. Still, psychology became one of the more popular degrees at historically black colleges and universities. For example, at Wilberforce College in Ohio, there was an honors psychology club with more than 40 members in 1914 (Guthrie, 1998).

Unlike the white institutions, black colleges and universities stressed the applied aspects of psychology. Consequently, applied fields like education and guidance counseling were more commonly taught. Students could use their training to return to their communities and offer practical services through the (separate) school systems and churches. In the first two generations after the discipline of psychology was begun (ca. 1879–1920), then, psychology was being taught and used by minority, primarily African American, communities. It is important to keep in mind that because the opportunities for higher education for minorities were very limited, many African American communities focused on sending a few bright students, what DuBois (1903) called the “talented tenth,” to college to be trained as teachers, ministers, lawyers, dentists, and doctors.

Early African American Psychologists

One of the first, if not the first, African American to work on psychological topics was Charles Henry Turner (1867–1923). Turner was part of the small group of animal behavior researchers working in the tradition of Charles Darwin at the end of the 19th century and into the first few decades of the 20th century. The research tradition developed by these individuals became the new field of comparative psychology. However, Turner trained as a biologist at the bachelor’s (1891) and master’s (1892) levels, then earned his PhD in 1907 in zoology. He spent his professional career as an educator at black institutions. His scientific work focused on adaptations (learning) in a variety of species: ants, cockroaches, spiders, pigeons, snakes, bees, wasps, and moths (Abramson, 2006). Turner had 70 scientific publications over the course of his 33-year career, yet he was never able to secure a faculty position at a research-oriented university. He was a finalist for a position at the University of Chicago, but he was rejected when a new chairman who did not want a person of color on his faculty was hired (DuBois, 1929).

At Clark University in Worcester, Massachusetts, there were several African American students who earned master’s or doctoral degrees between 1915 and 1920. The aforementioned G. Stanley Hall, founder and president of Clark, believed that Africans and people of African descent were in the adolescent stage of civilization development. In Hall’s (1904) view, African Americans were not inferior to whites, they were just not as developed and so could not be expected to perform equally the intellectual and educational tasks that were natural to whites. Of the black students there in this period, three earned degrees in psychology: Howard Hale Long (MA, 1916), J. Henry Alston (MA, 1920), and Francis Sumner (PhD, 1920). All had important careers in psychology.

Long was a professor at historically black colleges in Georgia and Tennessee before moving to a position as Associate Superintendent for Research in the Washington, DC, public school system (Guthrie, 1998). He was a major contributor to the psychological and educational research literature on intelligence, academic achievement, and race in his career. His 1935 article in the Howard University *Journal of Negro Education*, "Some psychogenic hazards of segregated education of Negroes," was an important contribution to the evidence used by the National Association for the Advancement of Colored People (NAACP) Legal Defense Fund as part of their legal efforts to end segregation in public schools. Those efforts eventually culminated successfully in the U.S. Supreme Court decision *Brown v. Board of Education*, handed down in 1954 (see the following for more on this).

Francis Sumner was Hall's last doctoral student. He earned his PhD from Clark University in 1920, with a thesis titled "Psychoanalysis of Freud and Adler." He was the first African American to earn a doctorate in psychology (Guthrie, 1998; Sawyer, 2000). As notable as this accomplishment was, Sumner's career contributions far exceeded this beginning. His first full-time appointment was at West Virginia Collegiate Institute (now West Virginia State College), where he served from 1921 to 1928. He then moved to Howard University in Washington, DC, where he remained until his death in 1954. At Howard, he led the efforts to create an independent Department of Psychology (1930). From this point until his death, Howard developed the most outstanding department of psychology in any historically black institution (Guthrie, 1998). Among the many who earned psychology degrees (bachelor's or master's) there during this time were Kenneth and Mamie Clark, Keturah Whitehurst, James Bayton, and Alonzo Davis. Each of these individuals made important contributions to psychology.

Sumner made research contributions as well with published articles on emotions, religion, education, and intelligence. He, along with other black psychologists of this period, contributed to the intense debate then raging about the relative influences of environment and heredity in psychological abilities (Dewsbury, 1984; Sumner, 1928).

From the mid-to-late 1930s into the 1950s and 1960s, the number of minority psychologists slowly but steadily increased. Herman G. Canady (PhD, Northwestern, 1941) replaced Francis Sumner at West Virginia State College and pursued a vigorous research program for many years. Canady's interest was in clinical problems, with a focus on issues such as the race of the examiner in testing children. In 1938, Canady organized psychologists in a special section of the all-black American Teachers Association (ATA). The all-white National Education Association would not admit minority educators. At the 1938 ATA meeting, Division 6, Department of Psychology, was formed. It was the first professional organization of psychologists of color. When APA reorganized during World War II, Canady led a delegation from Division 6 to participate in the reorganization (Capshew & Hilgard, 1992).

In this period, ethnic minority psychologists were gaining strength, even if their numbers were not high. There was great vitality in the psychology departments at Howard, West Virginia State, and other centers. At the same time, there was a great cultural expansion in Harlem that encompassed music, painting, poetry, and literary novels. This was the era of Duke Ellington, Richard Wright, Zora Neale Hurston (an anthropologist as well as a novelist), Langston Hughes, and Ralph Ellison. Black communities across America knew what was happening in Harlem and took special pride in the accomplishments springing from this cultural and artistic center.

In the 1930s, African American women began earning doctorates in psychology. Inez Beverly Prosser was the first, earning a PhD in psychology from Cincinnati in 1933. She had already had an outstanding career in education in institutions in Texas and Mississippi. A fellowship from the Rockefeller Foundation's General Education Board while she was a professor at Tougaloo College in Mississippi made it possible for her to attend the University of Cincinnati for her doctoral work (Benjamin, Henry, & McMahon, 2005). Cincinnati, though it had a troubled past of racist actions by white citizens, was in a period of relative calm in the early 1930s. In fact, the university had a legacy of offering higher education to black students, at least for some of its degree programs (Benjamin, Henry, & McMahon, 2005). Charles Henry Turner had, in fact, earned his bachelor's and master's degrees in biology there (Abramson, 2006). With a bright future ahead of her, Prosser proudly received her doctorate in 1933. Unfortunately, she was killed in an automobile accident in 1934, cutting short what had already been a brilliant career (Guthrie, 1998).

Other women who received their doctorates in this period included Keturah Whitehurst (PhD, Radcliffe, 1952), who taught at Meharry Medical College (Nashville) in the 1950s before moving to Virginia State College in 1958, where she spent the remainder of her career and where she "mothered" several leaders of black psychology, including Aubrey Perry (Farley & Perry, 2002). Ruth Howard (PhD, University of Minnesota, 1934) was a developmental psychologist, having worked with the noted developmentalist Florence Goodenough. Howard later married the African American psychologist Albert Beckham, who was the third African American to earn his doctorate in psychology (NYU, 1930). Alberta Banner Turner (PhD, Ohio State, 1937), was a clinical psychologist who served for many years with the Ohio Bureau of Juvenile Research (Guthrie, 1998). These women, and numerous others, forged rewarding careers despite the constant discrimination and lower pay they experienced as women of color.

The Clarks' Contributions

Mamie Phipps Clark (PhD, Columbia, 1944) came to psychology after an initial interest in pursuing a mathematics degree at Howard University. She met Kenneth Bancroft Clark when both were students at Howard, and he persuaded her to change her major to psychology (Guthrie, 1990). Both she and Kenneth earned their bachelor's and master's degrees from Howard before moving on to Columbia University. (Kenneth, a few years ahead of Mamie, earned his doctorate at Columbia in 1940.) It was at Howard that Mamie Phipps, having secretly married Kenneth in 1937, began her work on racial identification in children. This work became a joint effort with Kenneth once she moved to New York. In their work with black children in both northern and southern U.S. settings, the Clarks found evidence that segregation inflicted psychological damage on the children (Phillips, 2000). Kenneth Clark later reported that the work had been terribly disturbing to both of them and they had great reservations about pursuing it. However, with the encouragement of the NAACP Legal Defense fund, Kenneth agreed to use the results of their studies in the Fund's court challenges to school segregation (Jackson, 2001).

It was this work along with other social science research on the effects of racial segregation and discrimination that was considered by the U.S. Supreme Court in their landmark decision *Brown v. Board of Education*, when the Court ruled in 1954 that segregation in public schools was unconstitutional (Pickren, 2004a). It was the first time that psychological research had been used by the Court in making their decision. Not unusual for the time, the predominantly white APA made no mention of it in any of their fora (Benjamin & Crouse, 2002). The decision and its defiance by school boards in both the North and the South were critical turning points in U.S. history. As it fit within the larger civil rights movement, the work of Kenneth and Mamie Clark that was used by the Court became one of the most important applications of psychology to the cause of social justice.

But the work of the Clarks was much more diverse than the "doll studies" that were part of the evidence presented to the Court. Because they both wanted to make a difference in the lives of children and families, the Clarks, with their own money, started the Northside Center for Child Development in Harlem in 1946. It was, for Kenneth, the "convergence of academic life and social policy" (as cited in Nyman, 1976, p. 111). He was the research director at the center from 1946 to 1952. For Mamie, it was a way for her to use her training, expertise, and sense of social justice to "give children security" (Lal, 2002).

The Clarks faced many challenges at the Northside Center (Markowitz & Rosner, 2000). For example, the Clarks discovered that the New York Board of Education had policies that facilitated the easy labeling of minority or immigrant children as mentally retarded or developmentally delayed and then shuffled them off to special classes. This has been a problem for well over a century in the U.S., and it continues today. This issue became one that the Clarks and their staff at Northside dealt with repeatedly.

Mamie, as director, soon discovered that focusing solely on the problems of individuals and their families was inadequate. There were structural problems—housing, organization of work, schools, social and educational policies—that could not be ignored if these children were to be helped. Kenneth Clark held strong views about the need to address the larger social structure, as he became well aware from his research that the provision of clinical services alone would not make an enduring difference in the health of the community (Nyman, 1976). The need to address the larger social problems and pathologies in the social system was at the crux of what the Clarks wanted to do.

In 1962, Kenneth became the director of the President's Committee on Juvenile Delinquency program in Harlem, the Harlem Youth Opportunities Unlimited. Initially, this effort to revive Harlem as a place of opportunity was successful. Kenneth and his staff set out to make a structural difference and improve individual lives. Through a jobs program for youth and community projects to freshen up neighborhoods and improve services, it seemed like there was hope when communities took action collectively.

However, other black leaders, mainly Adam Clayton Powell, then the representative to the U.S. Congress for that district, betrayed Clark over money. Clark became disillusioned about the possibility of real change. In his disillusionment he wrote *Dark Ghetto* (1965). It reflected his despair over structural and individual racism.

In 1968, Kenneth Clark was asked by the leaders of APA to allow his name to be placed on the ballot for the presidency of APA. The association was, at this time, struggling with how to respond to the massive social changes and social problems then facing the nation and shaking up APA. The new, militant Association of Black Psychologists had laid challenges of racism and irrelevance at APA's door, and APA was not sure how to respond (Pickren, 2004b).

Clark became APA president-elect in 1969 and served his term as president from September 1970 through August 1971. Among his enduring contributions were his actions to establish internal governance structures within APA to deal with issues of social justice and public interest. The APA membership approved the creation of the Board for Social and Ethical Responsibility in Psychology (BSERP). This board, which no longer exists, inaugurated a sea change in APA's approach to psychology, making APA much more receptive to issues of race, ethnicity, sexual orientation, and disability through such structures as the Board of Ethnic Minority Affairs and the current Board for the Advancement of Psychology in the Public Interest (Pickren & Tomes, 2002).

The Association of Black Psychologists

Pressuring APA from another side was a group of African American psychologists, alluded to previously. These psychologists were mostly of a younger generation than the Clarks and were inspired by the more militant views of Elijah Muhammed, Malcolm X, and Frantz Fanon. The Clarks were committed to integration and the abolition of the color line. The younger black psychologists were committed to black identity and black nationalism. Rather than focusing on the damage to black communities and accepting victimization, these psychologists portrayed black communities as sources of strength and black families as sources of resilience (White, 1972). This was at the time that the national discourse, led by such well-meaning whites as Daniel Patrick Moynihan, was labeling black families as "tangles of pathology."

What upset many younger and older black APA members was the lack of action on APA's part to address institutional racism and the failure of psychology graduate programs to recruit and retain minority students or train their students in theory and method relevant to social problems. In short, what many minority psychologists (accurately) perceived was that APA represented a white middle-class ideology committed to maintaining the status quo. And the status quo included using psychological tests to label and sort minority children and students as inferior or deficient, or, at best, culturally deprived. In 1962 the APA had created a committee to investigate the low numbers of minority psychologists, but the committee was bogged down and had done nothing by the late 1960s. Fed up with what they perceived as foot-dragging, a group of young black psychologists founded the Association of Black Psychologists (ABPsi) in 1968 at the APA meeting in San Francisco (B. H. Williams, 1997; R. Williams, 1974).

ABPsi was led by a dedicated group of mostly male psychologists. Some of the leaders were Henry Tomes, Robert Williams, Reginald Jones, Joseph White, Robert Green, Charles Thomas, Asa Hilliard, and others. In retrospect, the initial impetus for the organization was reactive; its members were angry and upset at the failures of what they perceived as a white-dominated psychology. However, these psychologists soon articulated a black psychology that focused on the strengths and resilience of black folks and black communities. In journals (e.g., *Black Scholar*, *Journal of Black Psychology*) and books (e.g., the several editions of Reginald Jones's *Black Psychology)*, White, Williams, Jones, Hilliard, Luther, X, and others articulated a positive psychology predicated on a worldview informed by the history and philosophy of people of African descent. One aspect that was present from the beginning and remains potent is the emphasis on community and how resolutions of problems come from reliance on community resources and the strength that the community gives to its members. This orientation stems from the communalism that is part of the cultural traditions of those of African, particularly West African, descent (Nobles, 1972). What the founders of ABPsi fostered and what has been maintained is a psychology of resilience and strength anchored in community.

ABPsi as an Organizational Model

Once the ABPsi was formed, it quickly became a role model for other organizations of ethnic minority psychologists in the United States. In 1970, Edward Casavantes, a Mexican American (Chicano) psychologist founded the first professional organization of Hispanic psychologists, the Association of Psychologists Por La Raza (APLR). At first, there were only a handful of members and the organization struggled to get recognition from APA (Pickren & Tomes, 2002). By the end of the 1970s, the group re-formed as the National Hispanic Psychological Association. The growth of the association was fostered by the financial support provided by the National Institute of Mental Health (NIMH) for conferences and training (Pickren, 2004b). In 1979 the *Hispanic Journal of Behavioral Science* was founded at the Spanish Speaking Mental Health Resource Center at UCLA, with Amado Padilla as the editor. After a period of lagging growth, the group was renamed the National Latino Psychological Association and began a period of growth at the beginning of the century, thus reflecting the rapid growth of the Latino/a population in the United States.

The Asian American Psychological Association (AAPH) was founded in 1972 by two brothers, Derald and Stanley Sue. By the late 1960s, there was growing unrest in the Asian community in the San Francisco Bay area regarding the inadequacy of mental health services. The brothers brought together leaders of various community groups, psychologists, social workers, and ministers who were involved in some way with mental health work. NIMH provided money to fund a meeting to sort the issues out, a meeting that became quite heated and confrontational (Pickren, 2004b). Out of this passionate meeting came a decision to more formally organize. As Stanley Sue later recounted, AAPA initially built its membership by going through the APA directory and contacting anyone who had an Asian-sounding name (Pickren & Nelson, 2007). From this small beginning, the AAPA has had a healthy growth pattern into the 21st century. It had a membership of over 400 by the year 2000.

The AAPA has been involved in a range of issues since its founding. It has provided key liaisons to the NIMH and other federal agencies to assist with the development of workable policies for training minority mental health providers and to foster cultural competence in the training of all psychologists. There has been substantial progress on developing a body of theoretical work that is reflective of Asian American cultural experiences in psychology by members of the AAPA, including clinical training and social research. Derald Sue and other leaders of the association were among the first advocates of guidelines for multicultural counseling. The *Journal of the Asian American Psychological Association* began in 1979. Dr. Richard Suinn (b. 1933) was elected to serve as the president of the APA in 1999, the first Asian American to serve in this office.

The Society of Indian Psychologists (SIP) grew out of two different organizational efforts in the early 1970s. Carolyn Attneave founded the Network of Indian Psychologists in 1971. Joseph Trimble formed the American Indian Interest Group, also in 1971. The two groups merged in 1973 and changed the name to Society of Indian Psychologists in 1975 (Trimble, 2000). The membership of SIP has remained small but very active. Because of a legacy of racism, which endures to the present, fostered by official U.S. government policies, the mental health needs of American Indians have been and remain great. Efforts to increase the number of American Indian mental health providers have met with some success. The most notable success has been the Indians into Psychology Doctoral Education (INDPSYDE) program begun by Arthur L. McDonald in the mid-1980s. In the early years of the 21st century, INDPSYDE programs were in place at a few colleges and universities in the West and in Alaska. As a result there has been a slow growth in the number of mental health professionals of American Indian or Alaska Native descent. Logan Wright, in 1986, became the first person of American Indian heritage to be elected APA president (Trimble & Clearing Sky, in press).

TOWARD INCLUSIVENESS IN 21ST-CENTURY PSYCHOLOGY

As we can see, progress has been made in U.S. psychology in terms of making diversity a part of the agenda. Still, the membership of ethnic minorities in APA and other general psychological organizations lags well behind the proportions in the U.S. population. And, still, there is significant resistance to making training programs in the professional fields of psychology—clinical, counseling, and school—truly committed to the necessity of training all the students to be culturally competent in the provision of mental health services.

Where can we look for a model of how to do this best? Surprisingly, it is in many of the professional training schools, many offering the Doctor of Psychology degree, that we can find successful models of training for diversity. At the best programs, the administration and faculty are characterized by the following:

- Commitment from all concerned parties
- Sufficient financial support: grants, fellowships, and so forth
- Flexible admission criteria: for example, life experiences
- Presence of one or more minority faculty members
- Nurturing and supportive environment
- Incorporation of cultural diversity into all aspects of program

Although these characteristics are critically important, they cannot be implemented in a cookie-cutter fashion. Rather, each program has to find a way to create a program that is sensitive and welcoming of diversity. Given the changing demographics of U.S. society, doing so is no longer an option if psychology is to remain a relevant discipline and profession. With successful implementation of these programs creating a vibrant pipeline for minority psychologists, we might optimistically predict ever-increasing diversity in the discipline itself. As *who psychologists are* changes, the psychological knowledge produced by the discipline will become increasingly relevant to all of its consumers.

REFERENCES AND FURTHER READINGS

Abramson, C. I. (2006). Charles Henry Turner, pioneer of comparative psychology. In D. A. Dewsbury, L. T. Benjamin, Jr., & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 6, pp. 36–49). Washington, DC: American Psychological Association.

Benjamin, L. T., Jr., & Crouse, E. M. (2002). The American Psychological Association's response to *Brown v Board of Education*: The case of Kenneth B. Clark. *American Psychologist, 57*, 38–50.

Benjamin, L. T., Jr., Henry, K. D., & McMahon, L. R. (2005). Inez Beverly Prosser and the education of African Americans. *Journal of the History of the Behavioral Sciences, 41*, 43–62.

Bohan, J. S. (Ed.). (1995). *Re-placing women in psychology: Readings toward a more inclusive history* (2nd ed.). Dubuque, IA: Kendall/Hunt Publishing Company.

Bond, M. H. (1997). *Working at the interface of cultures: Eighteen lives in social science.* London: Routledge.

Boring, E. G. (1951). The woman problem. *American Psychologist, 6,* 679–682.

Boring, E. G. (1961). *Psychologist at large.* New York: Basic Books.

Brock, A. C. (Ed.). (2006). *Internationalizing the history of psychology.* New York: New York University Press.

Bryan, A. I. (1986). A participant's view of the National Council of Women Psychologists: Comment on Capshew and Laszlo. *Journal of Social Issues, 42,* 181–184.

Bryan, A. I., & Boring, E. G. (1944). Women in American psychology: Prolegomenon. *Psychological Bulletin, 41,* 447–454.

Bryan, A. I., & Boring, E. G. (1946). Women in American psychology: Statistics from the OPP questionnaire. *American Psychologist, 1,* 71–79.

Bryan, A. I., & Boring, E. G. (1947). Women in American psychology: Factors affecting their professional careers. *American Psychologist, 2,* 3–20.

Capshew, J. H., & Hilgard, E. R. (1992). The power of service: World War II and professional reform in the American Psychological Association. In R. B. Evans, V. S. Sexton, & T. C. Cadwallader (Eds.), *The American Psychological Association: A historical perspective* (pp. 149–175). Washington, DC: American Psychological Association.

Capshew, J. H., & Laszlo, A. C. (1986). "We would not take no for an answer": Women psychologists and gender politics during World War II. *Journal of Social Issues, 42,* 157–180.

Chesler, P. (1972). *Women and madness.* Garden City, NY: Doubleday.

Comas-Diaz, L. (1991). Feminism and diversity in psychology: The case of women of color. *Psychology of Women Quarterly, 15,* 597–610.

Dewsbury, D. A. (1984). Comparative psychology in the twentieth century. Stroudsburg, PA: Hutchinson Ross.

DuBois, W. E. B. (1903). *The souls of black folks, essays and sketches*. Chicago: A. C. McClurg & Co.

DuBois, W. E. B. (1929, June). Postscript. *The Crisis,* 203–204, 212.

Farley, F., & Perry, A. M. (2002). Keturah Elizabeth Whitehurst. *American Psychologist, 57.*

Fernberger, S. (1932). The American Psychological Association: A historical summary, 1892–1930. *Psychological Bulletin, 29*, 1–89.

Furumoto, L. (1992). Joining separate spheres—Christine Ladd-Franklin, woman-scientist (1847–1930). *American Psychologist, 47,* 175–182.

Guthrie, R. V. (1998). *Even the rat was white: A historical view of psychology* (2nd ed.). Boston: Allyn & Bacon.

Hall, G. S. (1904). *Adolescence, its psychology and its relations to physiology, anthropology, sociology, sex, crime, religion and education* (2 vols.). New York: Appleton.

Henley, N. M. (1977). *Body politics: Power, sex, and nonverbal communication.* Englewood Cliffs, NJ: Prentice Hall.

Holliday, B. G., & Holmes, A. (2003). A tale of challenge and change: A history of ethnic minorities in psychology in the United States. In G. Bernal, J. E. Trimble, A. K. Berlew, & F. T. L. Leong (Eds.), *Handbook of racial and ethnic minority psychology* (pp. 15–64). Thousand Oaks, CA: Sage.

Jackson, J. P. (2001). *Social scientists for social justice: Making the case against segregation*. New York: New York University Press.

Kim, U., & Berry, J. W. (1993). *Indigenous psychologies: Research and experience in cultural context.* Newbury Park, CA: Sage.

Klein, A. G. (2002). *A forgotten voice: A biography of Leta Stetter Hollingworth.* Scottsdale, AZ: Great Potential Press.

Lal, S. (2002). Giving children security: Mamie Phipps Clark and the racialization of child psychology. *American Psychologist, 57*, 20–28.

Leong, F. (1995). History of Asian American psychology. *AAPA Monograph, 1,* 1–54.

Long, H. H. (1935). Some psychogenic hazards of segregated education of Negroes. *Journal of Negro Education, 4*, 336–350.

Louw, J. (2002). Psychology, history and society. *South African Journal of Psychology, 32*, 1–8.

Markowitz, G., & Rosner, D. (2000). *Children, race, and power.* New York: Routledge.

Mednick, M. T., & Urbanski, L. L. (1991). The origins and activities of APA's Division of the Psychology of Women. *Psychology of Women Quarterly, 15,* 651–663.

Milar, K. S. (1999). "A Coarse and Clumsy Tool": Helen Thompson Woolley and the Cincinnati Vocational Bureau. *History of Psychology*, *2*(3), 219–235.

Morawski, J. G. (1990). Toward the unimagined: Feminism and epistemology in psychology. In R. T. Hare-Mustin and J. Marecek (Eds.), *Making a difference: Psychology and the construction of gender* (pp. 150–183). New Haven, CT: Yale University Press.

Nobles, W. W. (1972). African philosophy: Foundations for a black psychology. In R. L. Jones (Ed.), *Black psychology* (pp. 18–32). New York: Harper & Row.

Nyman, L. (1976). Recollections: An oral history of the psychology department of the City College of the City University of New York. New York: Author.

O'Connell, A. N. (2001). *Models of achievement: Reflections of eminent women in psychology* (Vol. 3). Mahwah, NJ: Erlbaum.

O'Connell, A. N., & Russo, N. F. (Eds.). (1983). *Models of achievement: Reflections of eminent women in psychology.* New York: Columbia University Press.

O'Connell, A. N., & Russo, N. F. (Eds.). (1988). *Models of achievement: Reflections of eminent women in psychology* (Vol. 2). Hillsdale, NJ: Erlbaum.

Paranjpe, A. C. (2002). Indigenous psychology in the post-colonial context: An historical perspective. *Psychology and Developing Societies, 14,* 27–44.

Phillips, L. (2000). Recontextualizing Kenneth B. Clark: An Afrocentric perspective on the paradoxical legacy of a model psychologist-activist. *History of Psychology, 3*, 142–167.

Pickren, W. E. (Ed.). (2004a). Fifty years on: *Brown v Board of Education* and American psychology, 1954–2004 [Special issue]. *American Psychologist, 59*, 493–556.

Pickren, W. E. (2004b). Between the cup of principle and the lip of practice: Ethnic minorities and psychology, 1966–1980. *History of Psychology, 7,* 45–64.

Pickren, W. E., & Nelson, M. (2007). *Making a more inclusive psychology: History and contemporary issues* [Documentary DVD]. Washington, DC: American Psychological Association.

Pickren, W. E., & Tomes, H. (2002). The legacy of Kenneth B. Clark to the APA: The Board of Social and Ethical Responsibility in Psychology. *American Psychologist, 57*, 51–59.

Richards, G. (1997). *'Race,' racism, and psychology: Towards a reflexive history*. London: Routledge.

Russo, N. F., & Dumont, B. A. (1997). A history of Division 35 (Psychology of Women): Origins, issues, activities, and future. In D. A. Dewsbury (Ed.), *Unification through division: Histories of the divisions of the American Psychological Association* (Vol. 2, pp. 211–238). Washington, DC: American Pyschological Association.

Sawyer, T. F. (2000). Francis Cecil Sumner: His views and influence on African American higher education. *History of Psychology, 3*, 122–141.

Scarborough, E., & Furumoto, L. (1987). *Untold lives: The first generation of American women psychologists*. New York: Columbia University Press.

Schwesinger, G. C. (1943). Wartime organizational activities of women psychologists: II. The National Council of Women Psychologists. *Journal of Consulting Psychology, 7,* 298–299.

Shields, S. A. (1975). Ms. Pilgrim's progress: The contributions of Leta Stetter Hollingworth to the psychology of women. *American Psychologist*, *30,* 852–857.

Sumner, F. C. (1928). Environic factors which prohibit creative scholarship among Negroes. *School and Society, 22,* 112–118.

Thompson, H. B. (1903). *The mental traits of sex: An experimental investigation of the normal mind in men and women.* Retrieved October 18, 2007, from http://psychclassics.yorku.ca/Thompson/

Tiefer, L. (1991). A brief history of the Association for Women in Psychology, 1969–1991. *Psychology of Women Quarterly, 15,* 635–650.

Trimble, J. E. (2000). American Indian psychology. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 1, pp. 139–144). Washington, DC: American Psychological Association.

Trimble, J. E., & Clearing Sky, M. (in press). An historical profile of American Indians and Alaska natives in psychology. *Cultural Diversity and Ethnic Minority Psychology, 13*.

Unger, R. K. (1998). *Resisting gender: Twenty-five years of feminist psychology.* London: Sage.

Walsh, M. R. (1985). Academic professional women organizing for change: The struggle in psychology. *Journal of Social Issues, 41*, 17–28.

Warren, W. (1999). *Black women scientists in the United States*. Bloomington: Indiana University Press.

Weisstein, N. (1971). Psychology constructs the female. *Journal of Social Education, 35*, 362–373.

White, J. L. (1972). Toward a black psychology. In R. L. Jones (Ed.), *Black psychology* (pp. 43–50). New York: Harper & Row.

Williams, B. H. (1997). Coming together: The founding of the Association of Black Psychologists. *Dissertation Abstracts International.* (UMI No. 9822882).

Williams, R. (1974). A history of the Association of Black Psychologists: Early formation and development. *Journal of Black Psychology, 1*, 9–24.

Women's Programs Office. (2003). *Women in the American Psychological Association*. Washington, DC: American Psychological Association. Retrieved September 27, 2006, from http://www.apa.org/pi/wpo/wapa03.pdf

5

Conducting Research on the History of Psychology

David B. Baker
University of Akron

Ludy T. Benjamin, Jr.
Texas A&M University

Every day, psychologists make history. It can be in an act as small as sending an e-mail or as large as winning a Nobel Prize. What remains of these acts and the contexts in which they occur are the data of history. When transformed by historians of psychology to produce narrative, these data represent our best attempts to make meaning of our science and profession.

The meaning that is derived from the data of history is most often made available to students of psychology through a course in the history of psychology. For a variety of reasons, the history of psychology has maintained a strong presence in the psychology curriculum at both the undergraduate and graduate levels for as long as there has been a psychology curriculum in America (Fuchs & Viney, 2002; Hilgard, Leary, & McGuire, 1991). As a result, most students will have some exposure to the subject matter and some sense of its importance.

Why are psychologists so interested in their own history? In trying to answer this question, consider the following quotations from two eminent British historians. One, Robin Collingwood (1946), wrote that the "proper object of historical study…is the human mind, or more properly the activities of the human mind" (p. 215). And the other, Edward H. Carr (1961), proposed that "the historian is not really interested in the unique, but what is general in the unique" and that "the study of history is a study of causes…the historian…continuously asks the question: Why?" (pp. 80, 113). Thus, according to these historians, to study history is to study the human mind, to be able to generalize beyond the characteristics of a single individual or single event to other individuals and other events, and to be able to answer the "why" of human behavior in terms of motivation, personality, past experience, expectations, and so forth. Historians are not satisfied, for example, with a mere description of the events of May 4, 1970, in which national guard troops killed four unarmed students on a college campus in Ohio. Description is useful, but it is not the scholarly end product that is sought. By itself, description is unlikely to answer the questions that historians want to answer. They want to understand an event, like the shootings at Kent State University, so completely that they can explain *why* it happened.

Collingwood (1946) has described history as "the science of human nature" (p. 206). In defining history in that way, Collingwood has usurped psychology's definition for itself. One can certainly argue about the scientific nature of history and thus his use of the term *science* in his definition. Whereas historians do not do experimental work, they are engaged in empirical work, and they approach their questions in much the same way that psychologists do, by generating hypotheses and then seeking evidence that will confirm or disconfirm those hypotheses. Thus the intellectual pursuits of the historian and the psychologist are not really very different. And so as psychologists or students of psychology, we are not moving very far from our own field of interest when we study the history of psychology.

Historians of psychology seek to understand the development of the discipline by examining the confluence of people, places, and events within larger social, economic, and political contexts. Over the last forty years the history of psychology has become a recognized area of research and scholarship in psychology. Improvements in the tools, methods, and training of historians of psychology have created a substantial body of research that contributes to conversations about our shared past, the meaning of our present divergence, and the promise of our future. In this chapter you will learn about the theory and practice of research on the history of psychology.

THEORY

Historiography refers to the philosophy and methods of doing history. Psychology is certainly guided by underlying philosophies and a diversity of research methods. A behaviorist, for example, has certain assumptions about the influence of previous experience, in terms of a history of punishment and reinforcement, on current behavior. And the methods of study take those assumptions into account in the design and conduct of experiments. A psychoanalytic psychologist, on the other hand, has a very different philosophy and methodology in investigating the questions of interest, for example, believing in the influence of unconscious motives and using techniques such as free association or analysis of latent dream content to understand those motives. Historical research is guided in the same way. It will help you understand history by knowing something about its philosophy and methods as well.

The historical point of view is highly compatible with our notions of our science. Psychologists tend to view individuals in developmental terms, and historians of psychology extend this point of view to encompass the developmental life of the discipline. Like any area of inquiry in psychology, historians of psychology modify their theories, principles, and practices with the accumulation of knowledge, the passage of time, and available technology. One simply needs to compare E. G. Boring's epic 1929 tome, *A History of Experimental Psychology*, with Duane and Sydney Ellen Schultz's 2004 text, *A History of Modern Psychology*, to see the difference that 75 years can make.

Approaches to history have changed dramatically over the last 75 years. Indeed much of the early research and scholarship in the history of psychology was ceremonial and celebratory. Most often it was not written by historians. It was, and in some circles remains, a reflexive view of history—great people cause great change. Such a view is naïve and simplistic. Psychological theories, research practices, and applications are all bound in a context, and it is this dynamic and fluid model that is the trend in historical research today. Just as inferential statistics have advanced from simple regression analysis to structural equation modeling, so too has historical research embraced a notion of multiple determinants and estimates of their relative impact on historical construction. In 1989 historian of psychology Laurel Furumoto christened this "the new history," a signifier denoting that historic research should strive to be more contextual and less internal.

Postmodern, deconstructionist, and social constructionist perspectives all share an emphasis on context, and have influenced historical research in psychology. The postmodern approach embraces a more critical and questioning attitude toward the enterprise of science (Anderson, 1998). The rise of science studies has led to what some have dubbed the "science wars" and to contentious arguments between those who see science as an honest attempt at objective and dispassionate fact-finding and those who see science (psychological and otherwise) as a political exercise subject to disorder, bias, control, and authority mongering. It is an issue that is present in today's history of psychology (for examples and discussions see Popplestone, 2004; Zammito, 2004).

Perhaps the largest growth in scholarship on the history of psychology has been in the area of intellectual history. As mentioned earlier, the construction of narrative in these works tends to eschew the older, more ceremonial, and internal histories in favor of a point of view that is more external and contextual. Rather than merely providing a combination of dates and achievements, modern historical scholarship in psychology tends to illuminate. The value of this point of view is in its contributions to our ongoing discussions of the meanings and directions of our field. The ever-expanding universe that psychology occupies and the ongoing debates of the unity of psychology are sufficient to warrant consideration and discussion of how our science and practice have evolved and developed. Historical analysis offers insight into personal, professional, and situational variables that impact and influence the field.

There is also a growing interest in what can be termed the material culture of psychology. The objects and artifacts that occupy psychological laboratories and aid our assessment of mind and behavior are becoming objects of study in their own right (Robinson, 2001; Sturm & Ash, 2005). For example, we continue to study reaction time and memory but we no longer use Hipp chronoscopes or mechanical memory drums. Changes in technology bring changes in methodologies and a host of other variables that are of interest to the historian of psychology.

Another area of increased interest and attention is the impact that racism and discrimination have had on the field. Traditionally underrepresented groups in psychology have often been made invisible by the historical record, but recent scholarship seeks to illuminate the people, places, and practices that have been part of both the problem and the solution to some of the 20th century's most vexing questions on race, gender, and religion (for examples see Philogène, 2004; Winston, 2004).

METHODS

Psychologists typically study contemporary events (behaviors and mental processes), whereas historians study events of the distant past. Both might be interested in the same behavior, but the time frame and the methods are usually distinct. Psychologists are interested in marriage, for example, and they might study marriage using surveys, ex post facto methods, or quasi-experimental designs using a sample of married couples (or perhaps divorced couples). Historians, on the other hand, would be likely to look at marriage, for example, as an institution in Victorian England, and they would be unable to use any of the methods listed previously as part of the arsenal of the psychologist. The questions on marriage that would interest psychologists and historians might be similar—how are mates selected in marriage, at what age do people marry, what roles do wives and husbands play in these marriages, what causes marriages to end? But again, the methods of research and the time frame for the events would be different.

History, then, is the branch of knowledge that attempts to analyze and explain events of the past. The explanatory product is a narrative of those events, a story. Central to telling any historical story is the accumulation of facts. We typically think of facts as some kind of demonstrable truth, some real event whose occurrence cannot be disputed. Yet facts are more elusive, as evidenced in the typical dictionary definition, which notes that a fact is information that is "presented" as objectively real. Historians present as fact, for example, that an atomic bomb was dropped on the Japanese city of Hiroshima on August 6, 1945. Because of detailed records of that event, as well as many eyewitness accounts, that fact seems indisputable; however, there are other kinds of facts.

In addition to the date of the bombing of Hiroshima, historians have also presented a number of facts relevant to the decision made by the United States government to drop that bomb. Not surprisingly, those facts are more debatable. Thus facts differ in terms of their certainty. Sometimes that is because evidence is incomplete and much inference has to be made, sometimes it is because evidence is contradictory, and sometimes it is because of bias introduced in the observation or in the interpretation of these events. Flawed though they may be, facts are the basis of history. It is the job of the historian to uncover these items of the past and to piece them together in an account that is as accurate as can be constructed.

In contemporary historiography, the researcher must always be alert to bias in the selection and interpretation of facts. Objectivity is a critical goal for the historian. Carr (1961) has argued that objectivity is indeed only a dream: "The emphasis on the role of the historian in the making of history tends, if pressed to its logical conclusion, to rule out any objective history at all: history is what the historian makes" (p. 29).

Like psychologists, historians are human too, and they bring to their task a bundle of prejudices, preconceptions, penchants, predispositions, premises, and predilections. Such baggage does not mean that they abandon their hope for objectivity, nor does it mean that their histories are hopelessly flawed. Good historians know their biases. They use their understanding of them to search for evidence in places where they might not otherwise look or to ask questions that they would not ordinarily ask. When this searching and questioning causes them to confront facts contrary to their own views, they must deal with those facts as they would with facts that are more consistent with their biases.

Bias in history begins at the beginning: "The historian displays a bias through the mere choice of a subject…" (Gilderhus, 1992, p. 80). There are an infinite number of historical subjects to pursue. The historian selects from among those, often selecting one of paramount personal interest. The search within that subject begins with a question or questions that the historian hopes to answer, and likely the historian starts with some definite ideas about the answers to those questions.

Bias is evident too in the data of history. It can occur in primary source material—for example, census records or other government documents—even though such sources are often regarded as quite accurate. Yet such sources are inherently biased by the philosophies underlying the construction of the instruments themselves and the ways in which those instruments are used. Secondary sources too are flawed. Their errors occur in transcription, translation, selection, and interpretation.

Oral histories are subject to the biases of the interviewer and the interviewee. Some questions are asked, while others are not. Some are answered, and others are avoided. And memories of events long past are often unreliable. Manuscript collections, the substance of modern archives, are selective and incomplete. They contain the documents that someone decided were worth saving, and they are devoid of those documents that were discarded or lost for a host of reasons, perhaps known only to the discarder.

After they have selected a topic of study and gathered the facts, historians must assemble them into a narrative that can also be subject to biases. Leahey (1986) reviews some of the pitfalls that modern historians of science want to avoid. These include Whig history, presentism, internalist history, and Great Man theories. Whig history refers to historical narrative that views history as a steady movement toward progress in an orderly fashion. Presentism is the tendency to view the past in terms of current values and beliefs. Internalist history focuses solely on developments within a field and fails to acknowledge the larger social, political, and economic contexts in which events and individual actions unfold. Great Man theories credit single, unique individuals (most often white males) as makers of history without regard for the impact that the spirit of the times (often referred to as the zeitgeist) has on the achievements of individuals. Avoiding these errors of interpretation calls for a different approach, which Stocking (1965) has labeled "historicism": an understanding of the past in its own context and for its own sake. Such an approach requires historians to immerse themselves in the context of the times they are studying.

These are just some of the hurdles that the historian faces in striving for objectivity. They are not described here to suggest that the historian's task is a hopeless one; instead, they are meant to show the forces against which historians must struggle in attempts at accuracy and objectivity. Carr (1961) has characterized the striving for this ideal as follows:

> When we call a historian objective, we mean, I think, two things. First of all, we mean that he has the capacity to rise above the limited vision of his own situation in society and in history.…Secondly, we mean that he has the capacity to project his vision into the future in such a way as to give him a more profound and lasting insight into the past than can be attained by those historians whose outlook is entirely bounded by their own immediate situation. (p. 163)

In summary, history is a product of selection and interpretation. Knowing that helps us understand why books are usually titled "A History…" and not "The History…." There are many histories of psychology, and it would be surprising to find any historians so arrogant as to presume that their individual narratives constituted "*The* History of Psychology."

History research is often like detective work: the search for one piece of evidence leads to the search for another and another. One has to follow all leads, some of which produce no useful information. When all of the leads have been exhausted, then you can analyze the facts to see if they are sufficient for telling the story. The leads or the data of history are most often found in original source material. The published record provides access to original source material through monographs and serials that are widely circulated and available in most academic libraries (including reference works such as indexes, encyclopedias, and handbooks). Hard-to-find and out-of-print material (newspapers, newsletters) are now much more easily available thanks to the proliferation of electronic resources. Too often valuable sources of information (obituaries, departmental histories and records, and oral histories) that are vital to maintaining the historical record are not always catalogued and indexed in ways that make them readily available and visible. The most important of all sources of data are archival repositories. Within such repositories one can find records of individuals (referred to as manuscript collections) and organizations (termed archival collections). Manuscript collections preserve and provide access to unique documents such as correspondence, lab notes, drafts of manuscripts, grant proposals, and case records. Archival collections of organizations contain materials such as membership records, minutes of meetings, convention programs, and the like. Archival repositories provide, in essence, the "inside story," free of editorial revision or censure and marked by the currency of time as opposed to suffering the losses and distortion of later recall. In much the same way, still images, film footage, and artifacts such as apparatus and instrumentation aid in the process of historical discovery.

There are literally thousands of collections of letters of individuals, most of them famous, but some not. And in those historically significant collections are millions of stories waiting to be told. Michael Hill (1993) has described the joys of archival research in this way:

> Archival work appears bookish and commonplace to the uninitiated, but this mundane simplicity is deceptive. It bears repeating that events and materials in archives are not always what they seem on the surface. There are perpetual surprises, intrigues, and apprehensions.…Suffice it to say that it is a rare treat to visit an archive, to hold in one's hand the priceless and irreplaceable documents of our unfolding human drama. Each new box of archival material presents opportunities for discovery as well as obligations to treat the subjects of your… research with candor, theoretical sophistication, and a sense of fair play. Each archival visit is a journey into an unknown realm that rewards its visitors with challenging puzzles and unexpected revelations. (pp. 6–7)

"Surprise, intrigue, apprehension, puzzles, and discovery"—those are characteristics of detective work, and historical research is very much about detective work.

The papers of important psychologists are spread among archives and libraries all over the world. In the United States you will find the papers of William James and B. F. Skinner in the collections at Harvard University. The papers of Hugo Münsterberg, a pioneer in the application of psychology to business, can be found at the Boston Public Library. The papers of Mary Whiton Calkins and Christine Ladd-Franklin, important early contributors to experimental psychology, can be found at Wellesley College and at Vassar College and Columbia University, respectively. The Library of Congress includes the papers of James McKeen Cattell and Kenneth B. Clark. Cattell was one of the founders of American psychology and a leader among American scientists in general, and Clark, an African American psychologist, earned fame when his research on self-esteem in black children was cited prominently in the U.S. Supreme Court decision that made school segregation illegal (*Brown v. Board of Education*, 1954).

The single largest collection of archival materials on psychology anywhere in the world can be found at the Archives of the History of American Psychology (AHAP) at the University of Akron in Akron, Ohio. Founded by psychologists John A. Popplestone and Marion White McPherson in 1965, its purpose is to collect and preserve the historical record of psychology in America (Baker, 2004). Central to this mission is the preservation of personal papers, artifacts, and media that tell the story of psychology in America. In archival terms, "papers" refers to one-of-a-kind (unique) items. Papers can include such things as correspondence (both personal and professional), lecture notes, diaries, and lab journals. Recently named a Smithsonian Affiliate, the AHAP houses more than 1,000 objects and artifacts that offer unique insights into the science and practice of psychology. Instruments from the brass-and-glass era of the late 19th century share space alongside such significant

20th century objects as the simulated shock generator used by Stanley Milgram in his famous studies of obedience and conformity, the flags of the Eagles and Rattlers of the Robbers Cave experiment by Muzafir and Carolyn Sherif, and the props that supported Phillip Zimbardo's well-known Stanford University prison studies.

Currently, the AHAP houses the personal papers of over 700 psychologists. There are papers of those representing experimental psychology (Leo and Dorothea Hurvich, Kenneth Spence, Ward Halstead, Mary Ainsworth, Frank Beach, Knight Dunlap, Dorothy Rethlingshafer, and Hans Lukas-Tuber), professional psychology (David Shakow, Edgar Doll, Leta Hollingworth, Herbert Freudenberger, Sidney Pressey, Joseph Zubin, Erika Fromm, Jack Bardon, Robert Waldrop, Marie Crissey, and Morris Viteles), and just about everything in between. Also included are the records of more than 50 psychological organizations, including the American Group Psychotherapy Association, the Association for Women in Psychology, Psi Chi, Psi Beta, the Association for Humanistic Psychology, the International Council of Psychologists, and the Psychonomic Society. State and regional association records that can be found at the AHAP include those of the Midwestern Psychological Association, the Ohio Psychological Association, and the Western Psychological Association. The test collection includes more than 8,000 tests and records. There are more than 15,000 photographs and 6,000 reels of film, including home movies of Freud, footage of Pavlov's research institute, and research film from Arnold Gesell and the Yale Child Study Center. All of these materials serve as trace elements of people, places, and events to which we no longer have access. These archival elements are less fallible than human memory, and if properly preserved, are available to all for review and interpretation. Because an in-person visit to the Archives of the History of American Psychology is not always possible, the AHAP is seeking to make more of its collection available online (www.uakron.edu/ahap). Indeed, with the advent of the information age, material that was once available only by visitation to an archival repository can now be scanned, digitized, and otherwise rendered into an electronic format. From the diaries and correspondence of women during the civil war to archival collections of animation movies, the digital movement is revolutionizing access to original source material. More information on electronic resources in the history of psychology can be found in the annotated bibliography at the end of this chapter.

All archives have a set of finding aids to help the researcher locate relevant materials. Some finding aids are more comprehensive than others. Finding aids are organized around a defined set of characteristics that typically include the following:

- Collection dates (date range of the material)
- Size of collection (expressed in linear feet)
- Provenance (place of origin of a collection, previous ownership)
- Access (if any part of the collection is restricted)
- Finding aid preparer name and date of preparation
- Biographical/historical note (a short, succinct note about the collection's creator)
- Scope and content note (general description and highlights of the collection)
- Series descriptions (headings used to organize records of a similar nature)
- Inventory (description and location of contents of a collection)

Even if an on-site review of the contents of a collection is not possible, reviewing finding aids can still be useful because of the wealth of information they provide.

APPLICATIONS

In the mid-1960s, a critical mass of sorts was achieved for those interested in teaching, research, and scholarship in the history of psychology. Within the span of a few years, two major organizations appeared: Cheiron: The International Society for the History of the Social and Behavioral Sciences, and Division 26 (Society for the History of Psychology) of the American Psychological Association (APA). Both sponsor annual meetings, and both are affiliated with scholarly journals (Cheiron is represented by the *Journal of the History of the Behavioral Sciences* and the Society for the History of Psychology by *History of Psychology*) that provide an outlet for original research. Two doctoral training programs in the history of psychology exist in North America. One is at York University in Toronto, Ontario, Canada, and the other is at the University of New Hampshire.

For most students in psychology, the closest encounter with historical research comes in the form of a project or paper as part of a requirement for a class on the history of psychology. Using the types of resources that we have described in this chapter, it should be possible to construct a narrative on any number of topical issues in psychology.

For example, the ascendancy of professional psychology with its concomitant focus on mental health is a topic of interest to historians of psychology and of considerable importance to many students who wish to pursue graduate training in professional psychology. Using archival materials, original published material, secondary sources, and government documents, a brief example of a historical narrative is provided.

World War II and the Rise of Professional Psychology

America's entrance into World War II greatly expanded the services that American psychologists offered, especially in the area of mental health. Rates of psychiatric illness among recruits were surprisingly high, the majority of discharges from service were for psychiatric reasons, and psychiatric casualties occupied over half of all beds in

Veterans Administration hospitals. Not only was this cause for concern among the military, it also alerted federal authorities to the issue among the general population. At the time, the available supply of trained personnel met a fraction of the need. In a response that was fast and sweeping, the federal government passed the National Mental Health Act of 1946, legislation that has been a major determinant in the growth of the mental health profession in America (Pickren & Schneider, 2004). The purpose of the act was clear:

> The improvement of the mental health of the people of the United States through the conducting of researches, investigations, experiments, and demonstrations relating to the cause, diagnosis, and treatment of psychiatric disorders; assisting and fostering such research activities by public and private agencies, and promoting the coordination of all such researches and activities and the useful application of their results; training personnel in matters relating to mental health; and developing, and assisting States in the use ofthe most effective methods of prevention, diagnosis, and treatment of psychiatric disorders. (Public Law 487, 1946, p. 421)

The act provided for a massive program of federal assistance to address research, training, and service in the identification, treatment, and prevention of mental illness. It created the National Institute of Mental Health (NIMH) and provided broad support to psychiatry, psychiatric social work, psychiatric nursing, and psychology for the training of mental health professionals (Rubenstein, 1975). Through the joint efforts of the United States Public Health Service and the Veterans Administration, funds were made available to psychology departments willing to train professional psychologists. Never before had such large sums of money been available to academic psychology. The grants and stipends available from the federal government allowed universities to hire clinical faculty to teach graduate students, whose education and training was often supported by generous stipends. It was these funds that subsidized the Boulder Conference on Graduate Education in Clinical Psychology in 1949 (Baker & Benjamin, 2000).

The chief architect of the Boulder model was David Shakow (1901–1981). At the time, there was no other person in American psychology who had more responsibility and influence in defining standards of training for clinical psychologists. In 1947, Shakow crafted a report on the training of doctoral students in clinical psychology that became the working document for the Boulder Conference of 1949 (APA, 1947; Benjamin & Baker, 2004; Felix, 1947).

By the 1950s, professional psychologists achieved identities that served their members, served their various publics, attracted students and faculty, and ensured survival by maintaining the mechanisms necessary for professional accreditation and later for certification and licensure. In the free-market economy, many trained for public service have found greener pastures in private practice.

The training model inaugurated by the NIMH in 1949 has continued unabated for five decades, planned and supported largely through the auspices of the American Psychological Association. The exigencies that called for the creation of a competent mental health work force have changed, yet the professional psychologist engineered at mid-century has endured, as has the uneasy alliance between science and practice.

This brief historical analysis shows how archival elements can be gathered from a host of sources and used to illuminate the contextual factors that contributed to a significant development in modern American psychology. This story could not be told without access to a number of original sources. For example, the inner workings of the two-week Boulder conference are told in the surviving papers of conference participants, including the personal papers of David Shakow that are located at Akron in the Archives of the History of American Psychology. Papers relevant to the Mental Health Act of 1946 can be found in the National Archives in Washington, DC. Information about the role of the Veterans Administration in contributing to the development of the profession of clinical psychology can be found in the oral history collection available at the archives of the APA. Such analysis also offers an opportunity for reflection and evaluation, and tells us some of the story of the bifurcation of science and practice that has resulted in American psychology. We believe that historical analysis provides a perspective that can contribute to our understanding of current debates and aid in the consideration of alternatives.

Indeed, almost any contemporary topic that a student of psychology is interested in has a history that can be traced. Topics in cognition, emotions, forensics, group therapy, parenting, sexuality, memory, and animal learning, to name but a very few, can be researched. Archival resources are often more readily available than most might think. Local and regional archives and university library special collections all are sources of original material. For example, students can do interesting research on the history of their own psychology departments (Benjamin, 1990). University archives can offer minutes of faculty meetings, personnel records (those that are public), college yearbooks (which often show faculty members, student groups, etc.), course catalogues, building plans, and many more items. Interviews can be conducted with retired faculty and department staff, and local newspapers can be researched for related stories. The work can be informative, instructive, and very enjoyable.

SUMMARY

In the end we are left with an important question: So what? What is the importance of the history of psychology? What do we gain? The history of psychology is not likely to serve as an empirically valid treatment for anxiety, nor is it likely to offer a model of how memory works. But that is not the point. It is easily argued that the history of psychology offers some instrumental benefits. The examination of psychology's past provides not only a more meaningful understanding of that past, but a more informed and enriched appreciation of our present, and the best crystal ball available in making predictions about our field's future. It aids critical thinking by providing a compendium of the trials, tribulations, and advances that accrue from the enormous questions we ask of our science and profession, and it offers the opportunity to reduce the interpersonal drift we seem to experience. In recent years, psychologists have become estranged from one another in ways that were unknown not all that long ago. Yet we share a connection, however tenuous, and it is found in our shared history.

At the risk of being labeled Whiggish, we would add that the history of psychology, professional and otherwise, has contributed to a corpus of knowledge that is real, tangible, and capable of improving the quality of life of all living things, including our planet. There are few secrets; we know how to encourage recycling, we understand effective ways of treating drug addiction, we have methods for alleviating some of the suffering of mental illness, we can provide tools to improve reading skills, we can design good foster homes—the list could get quite long.

Our knowledge is a powerful tool that has developed over time and is a narrative worth knowing. Like any good story, it has its heroes and its villains, it is set in a time and place, and it offers us a message we can all hear and use.

REFERENCES AND FURTHER READINGS

American Psychological Association, Committee on Training in Clinical Psychology. (1947). Recommended graduate training program in clinical psychology. *American Psychologist, 2*, 539–558.

Anderson, P. (1998). *The origins of postmodernity*. London: Verso.

Archives of the History of American Psychology. (n.d.). Retrieved on October 18, 2007, from http://www.uakron.edu/ahap

Baker, D. B. (2004). Thirty-five years of archival achievement. In D. Baker (Ed.), *Thick description and fine texture: Studies in the history of psychology* (pp. vii–x). Akron, OH: The University of Akron Press.

Baker, D. B., & Benjamin, L. T., Jr. (2000). The affirmation of the scientist-practitioner: A look back at Boulder. *American Psychologist, 55,* 241–247.

Benjamin, L. T., Jr. (1990). Involving students and faculty in preparing a departmental history. *Teaching of Psychology, 17,* 97–100.

Benjamin, L. T., Jr. (2006). *A history of psychology in letters* (2nd ed.). Malden, MA: Blackwell.

Benjamin, L. T., Jr., & Baker, D. B. (2004). *From séance to science: A history of the profession of psychology in America*. Belmont, CA: Thomson/Wadsworth.

Boring, E. G. (1929). *A history of experimental psychology*. New York: Century Co.

Carr, E. H. (1961). *What is history?* New York: Random House.

Collingwood, R. G. (1946). *The idea of history*. London: Oxford University Press.

Fancher, R. E. (2004). A tale of two institutions: York University's history of psychology program and the Archives of the History of American Psychology. In D. Baker (Ed.), *Thick description and fine texture: Studies in the history of psychology* (pp. 162–173). Akron, OH: The University of Akron Press.

Felix, R. H. (1947, December 12). *Memo to National Advisory Health Council*. (Shakow Papers, M1375). Akron, OH: The University of Akron, Archives of the History of American Psychology.

Freedheim, D. K. (Ed.). (2003). *History of psychology.* In I. B. Weiner (Series Ed.), *Handbook of psychology* (Vol. 1, pp. 492–495). New York: Wiley.

Fuchs, A. H., & Viney, W. (2002). The course in the history of psychology: Present status and future concerns. *History of Psychology, 5,* 3–15.

Furumoto, L. (1989). The new history of psychology. In I. S. Cohen (Ed.), *G. Stanley Hall lecture series* (Vol. 9, pp. 5–34). Washington, DC: American Psychological Association.

Gilderhus, M. T. (1992). *History and historians: A historiographical introduction* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Green, C. D. (n.d.). *Classics in the history of psychology.* Retrieved on October 18, 2007, from http://psychclassics.yorku.ca/

Green, C. D. (n.d.). *This week in the history of psychology.* Retrieved on October 18, 2007, from http://psychclassics.yorku.ca/christo/podcasts/

Guthrie, R. V. (2003) *Even the rat was white: A historical view of psychology* (2nd ed.). Boston: Allyn & Bacon.

Hilgard, E. R., Leary, D. E., & McGuire, G. R. (1991). The history of psychology: A survey and critical assessment. *Annual Review of Psychology, 42,* 79–107.

Hill, M. R. (1993). *Archival strategies and techniques.* Newbury Park, CA: Sage.

Leahey, T. H. (1986). History without the past. *Contemporary Psychology, 31,* 648–650.

Philogène, G. (Ed.). (2004). *Racial identity in context: The legacy of Kenneth B. Clark*. Washington, DC: American Psychological Association.

Pickren, W. E., & Dewsbury, D. A. (Eds.). (2002). *Evolving perspectives on the history of psychology*. Washington, DC: American Psychological Association.

Pickren, W. E., & Schneider, S. F. (2004). *Psychology and the National Institute of Mental Health: A historical analysis of science, practice, and policy*. Washington, DC: American Psychological Association.

Popplestone, J. A. (2004). Reinventing the past through interpretation: Reflections on the history of psychology—35 years in the trenches. In T. C. Dalton & R. B. Evans (Eds.), *The life cycle of psychological ideas: Understanding prominence and the dynamics of intellectual change* (pp. 59–81). New York: Kluwer.

Robinson, D. K. (2001). Reaction-time experiments in Wundt's institute and beyond. In R. Rieber & D. K. Robinson (Eds.), *Wilhelm Wundt in history: The making of a scientific psychology* (pp. 161–204). New York: Kluwer.

Rubenstein, E. A. (1975). Unpublished interview with Robert Felix. Available from the archives of the American Psychological Association.

Schultz, D. P., & Schultz, S. E. (2004). *A history of modern psychology* (8th ed.). Belmont, CA: Wadsworth/Thompson.

Stocking, G. W., Jr. (1965). On the limits of 'presentism' and 'historicism' in the historiography of the behavioral sciences. *Journal of the History of the Behavioral Sciences, 1,* 211–218.

Sturm, T., & Ash, M. (2005). Roles of instruments in psychological research. *History of Psychology, 8,* 3–34.

Winston, A. S. (Ed.). (2004). *Defining difference: Race and racism in the history of psychology.* Washington, DC: American Psychological Association.

Zammito, J. H. (2004). *A nice derangement of epistemes: Post-positivism in the study of science from Quine to Latour.* Chicago: University of Chicago.

PART II

Basic Research Methodology and Analytic Techniques

6

Statistical Techniques and Analysis

Chris Spatz

Hendrix College

Statistics? What does statistics have to do with psychology?

Perhaps you have heard this question. It is a fairly common one and usually reflects the belief that psychologists are people who help others with emotional problems, mostly by talking with them. Why would such a person need to know about statistics? Of course, as this two-volume work shows, psychologists have many other interests and do many other things besides helping people with emotional problems. Nevertheless, even if psychologists restricted themselves to this one endeavor, statistics would be a necessary component of their field because the particular techniques that clinical and counseling psychologists use were selected after being tested and compared with other techniques in experiments that were are analyzed with statistics.

Of course, because psychologists are interested in memory, perception, development, social relations, and many other topics as well as emotional problems, they design and conduct experiments to answer all kinds of questions. For example,

- What events that occur after an experience affect the memory of the experience?
- Do people experience visual illusions to different degrees?
- In children, do the number of neural connections decrease from age 12 months to age 36 months?
- Do people work just as hard if an additional member joins their team?

To answer questions such as these, psychologists who study memory, perception, development, and social relations design experiments that produce data. To understand the data and answer the questions, statistical analyses are required. Thus, statistics courses are part of the college education of most every psychology major. In addition to being necessary for answering psychological questions, statistical analyses are used in a variety of other fields and disciplines as well. Actually, any discipline that poses comparative questions and gathers quantitative data can use statistics.

DEFINITIONS

Statistics and Parameters

Statistics has two separate meanings. A number or graph based on data from a sample is called a statistic. In addition, *statistics* refers to a set of mathematical techniques used to analyze data. Thus, the mean and standard deviation of a sample are statistics, and *t* tests and chi square tests are statistics as well.

A number or a graphic that characterizes a population is called a *parameter*. Parameters are usually symbolized with Greek letters such as μ (mean) and σ (standard deviation). Statistics are usually symbolized with Latin letters such as *M* or $\bar{X}$ (mean) and *SD* or $\hat{s}$ (standard deviation).

Populations and Samples

The goal of researchers is to know about populations, which consist of all the measurements of a specified group. Researchers define the population. A population might consist of the size of the perceived error of all people who look at a visual illusion; it might be the work scores of all people who have an additional member join their team. Unfortunately, for most questions, it is impossible or impractical to obtain all the scores in a population. Thus, direct calculation of μ or σ is usually not possible.

The statistical solution to this problem is to use a sample as a substitute for the population. A sample is a subset of the population. Calculations (such as *M*) based on sample data are used to estimate parameters (such as μ). In actual practice, researchers are interested not in an absolute value of μ but in the relationships among μ's from different populations. However, basing conclusions about μ's on sample data introduces a problem because there is always uncertainty about the degree of the match between samples and their populations. Fortunately, inferential statistical techniques can measure the uncertainty, which allows researchers to state both their conclusion about the populations and the uncertainty associated with the conclusion.

Descriptive Statistics and Inferential Statistics Techniques

Descriptive statistics are numbers or graphs that summarize or describe data from a sample. The purpose of descriptive statistics is to reveal a particular characteristic of a sample or a score with just one or two numbers or a graph. Table 6.1 is a catalog of commonly used descriptive statistics for one variable.

Researchers often gather data from two distributions to determine whether scores in one distribution are related to scores in the other distribution (correlation) or to determine the degree of difference between the two distributions (effect size index). Table 6.2 is a catalog of commonly used descriptive statistics for two variables.

Inferential statistical techniques provide quantitative measures of the uncertainty that accompanies conclusions about populations that are based on sample data. Common examples of inferential statistical techniques include chi square tests, *t* tests, analysis of variance (ANOVA), and nonparametric tests.

Table 6.1 Descriptive statistics for one variable

Central tendency	**indicate a typical or representative score**
Mean	the sum of the scores divided by the total scores
Median	a point that divides scores into upper and lower halves
Mode	the most frequently occurring score
Variability	**spread of the scores**
Range	the highest score minus the lowest score
Interquartile range	two scores between which 50 percent of the distribution lies
Standard deviation	square root of the average of the squared deviations from the mean
Variance	average of the squared deviations from the mean
Form	**shape of the distribution**
Symmetrical	left and right sides are mirror images
Normal	conforms to the formula for the normal curve
Skew	degree of asymmetry of a distribution
Kurtosis	proportion of the distribution that is in the tails
Position	**individual score's position among all the scores**
Percentile	percent of distribution that is equal to or below this score
z score	number of standard deviations from the mean
T scores	standardized scores; mean = 50, standard deviation = 10
Outlier	extreme score separated from the others
Frequency distribution	**table of all scores and frequency of their occurrence**
Simple	scores arranged highest to lowest with frequency of occurrence of each score
Grouped	scores compiled into equal-size intervals with frequency of scores in each interval
Frequency displays	**figures that show all the scores**
Frequency polygon	display of dots connected with lines
Histogram	display of touching vertical rectangles
Bar graph	display of separated vertical rectangles
Stem and leaf	shows scores in the middle and their frequencies on the sides
Boxplot	shows at least the range, 25th, 50th, and 75th percentiles
Contingency table	an R x C table of two variables

Sampling Distributions and Degrees of Freedom

A *sampling distribution* (also called a probability density function) shows the results of repeated random sampling from the same population. For a particular sampling distribution, each sample is the same size and the same statistic is calculated on each sample. If the statistic calculated is *M*, then the sampling distribution (in this case, a sampling distribution of the mean) allows you to find the probability that any particular sample mean came from the population. Sampling distributions provide the probabilities that are the basis of all inferential statistical techniques.

One attribute of every sampling distribution is its *degrees of freedom*, which is a mathematical characteristic of sampling distributions. Knowing the degrees of freedom (*df*) for a particular data set is

Table 6.2 Descriptive statistics for two variables

Correlation	**Degree of relationship between variables**
Pearson	relationship of linear, scaled variables
Spearman	relationship of ranked variables
point biserial	relationship of a scaled and a dichotomous variable
η (eta)	relationship of curvilinear variables
Effect size index	**Amount of difference between two variables**
d	difference in two means per standard deviation
r	amount of variance two scaled variables have in common
point biserial *r*	amount of variance in common of a scaled and dichotomous variable
φ(phi)	degree of dependence in a 2 x 2 frequency table
ω^2 (omega squared)	variance in common of scaled variables
percent difference	difference in two variables using one as the standard
odds ratio	compares likelihood of two conditions from a 2 x 2 table of frequencies
Confidence interval	limits that capture a parameter with specified degree of confidence

necessary for choosing an appropriate sampling distribution. Formulas for *df* are based on sample size and restrictions that a model imposes on the analysis.

A BRIEF HISTORY OF INFERENTIAL STATISTICAL TECHNIQUES

An examination of scientific reports before 1900 that include statistical analyses reveals a few familiar descriptive statistics such as the mean, median, standard deviation, and correlation coefficient but none of today's inferential statistical techniques. In 1900, however, Karl Pearson introduced the chi square goodness-of-fit test. This test objectively assessed the degree of fit between observed frequency counts and counts expected based on a theory. Using the test, researchers could determine the probability (*p*) of the data they observed, if the theory was correct. A small value of *p* meant that the fit was poor, which cast doubt on the adequacy of the theory. Although the logic of this approach was in use before 1900, the chi square test was the first inferential statistical technique whose sampling distribution was based on mathematical assumptions that researchers could reasonably be assured were fulfilled. Chi square tests today are a widely used tool among researchers in many fields. The path that Pearson used to establish chi square tests revealed the way for others whose data were statistics other than frequency counts. That is, for those who wanted to use this logic to determine the probability of sample statistics such as means, correlation coefficients, or ranks, the task was to determine the sampling distribution of those statistics.

Of those efforts that are today part of the undergraduate statistics curriculum, the second to be developed was Student's *t* test. William S. Gosset was an employee of Arthur Guinness, Son & Company, where he applied his knowledge of chemistry and statistics to the brewing process. From 1906 to 1907, he spent several months in London working with Pearson. In 1908, using the pseudonym "Student," he published sampling distributions of means calculated from small samples. The result was that experiments that had one independent variable (IV) with two levels could be analyzed with a test that gave the probability of the observed difference, if it were the case that the two samples were from the same population. To derive these sampling distributions, Student (1908) assumed certain characteristics about the population. One of these was that the population was normally distributed.

In 1925 Ronald A. Fisher published *Statistical Methods for Research Workers*, a handbook that included summaries of his earlier work. This example-based instruction manual showed how to analyze data from experiments with more than two levels of the IV (one-way ANOVA) and experiments with more than one IV (factorial ANOVA). In addition, it showed how to remove the effects of unwanted extraneous variables from the data (analysis of covariance). Fisher's (1925) book was incredibly influential; 14 editions were published, and it was translated into six other languages.

Perhaps more important than Fisher's description of a general approach to analyzing data, regardless of the kind of experiment the data came from, was his solution to the problem of extraneous variables in experiments. In a simple but ideal experiment, there is an IV with two levels and one dependent variable (DV; see Chapter 9). The researcher's hope is to show that changes in the levels of the IV produce predictable changes in the DV. The most serious obstacle to establishing such cause-and-effect conclusions is extraneous variables, which are variables other than the IV that can affect the DV. In the worst case, an extraneous variable is confounded with the IV so that there is no way to know whether changes in the DV are due to the IV or the confounded extraneous variable. Fisher's designs either balanced the effects of extraneous variables over the different levels of the IV such that there was no differential effect on the DV or they permitted the effects of the extraneous variables to be removed from the analysis. It is not too strong to say that Fisher's methods revolutionized science in agronomy, biology, psychology, political science, sociology, zoology, and the other fields that rely on statistical analyses.

The logic of the approach of that Fisher and Gosset used was to tentatively assume that the samples observed came from populations that were identical (a hypothesis of no difference). A sampling distribution based on a hypothesis of no difference shows the probability of all samples from the population, including the one actually observed. If the probability of the actually observed sample is small, then support for the tentative hypothesis is weak. Following this reasoning, the researcher should reject the tentative hypothesis of no difference and conclude that the samples came from populations that were not identical.

What is the dividing point that separates small probabilities from the rest? Fisher's (1925) answer in *Statistical Methods for Research Workers* was to mention a value of $p = 0.05$ and continue with "…it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not" (p. 47). However, Fisher's own practice, according to Salsburg (2001), was to recognize three possible outcomes of an experiment: small p, large p, and intermediate p. A small p such as ≤ 0.05 led to rejection of the hypothesis of no difference. A large p such as .20 or greater led to the conclusion that if the treatments actually made a difference, it was a small one. An intermediate case of $0.05 < p < .20$ led to the conclusion that more data were needed (Salsburg, 2001). Remember that the accuracy of these probability figures is guaranteed only if certain assumptions about the populations are true.

In 1933 Jerzy Neyman and Egon Pearson (son of Karl Pearson) proposed revisions and extensions to Fisher's approach. They noted that in addition to the hypothesis of no difference in the populations, there is an unemphasized hypothesis—there is a difference in the populations. The follow-up to this seemingly trivial point led to developments such as Type I errors, Type II errors, and power, which are explained below.

Neyman and Pearson referred to the hypothesis of *no difference* as the null hypothesis and the hypothesis of *a difference* as the alternative hypothesis. Taken together, these two are exhaustive (they cover all possibilities) and mutually exclusive (only one can be true). Noting that there are two hypotheses revealed clearly that an experiment was subject to two kinds of errors. Rejecting the null hypothesis when it was true (due to an unlucky accumulation of chance) became known as a Type I error, an error that was clearly recognized by Fisher. However, failure to reject a false null hypothesis is also an error, and Neyman and Pearson called this an error of the second type (a Type II error). Identifying and naming the Type II error pushed statistical practice toward a forced decision between two alternatives: *reject the null hypothesis* or *fail to reject it*. In addition, recognizing the Type II error, which can occur only when the null hypothesis is false, led the way to an analysis of statistical power, which is the probability of rejecting a false null hypothesis.

In 1934 Neyman introduced the confidence interval (Salsburg, 2001). Calculated from sample data and utilizing a sampling distribution, a confidence interval is a range of values with a lower limit and an upper limit that is expected to contain a population parameter. The expectation is tempered by the amount of confidence the sampling distribution provides. Researchers most commonly choose a 95 percent level of confidence, but 90 percent and 99 percent are not uncommon.

In the 1940s another group of sampling distributions were derived so that probabilities could be determined for data that consisted of ranks rather than the more common, relatively continuous measures such as distance, time, and test scores. Frank Wilcoxon (1945) and Henry B. Mann and D. Ransom Whitney (1947) published tests that are appropriate for ranks. These tests were different from previous ones in that the populations sampled are not normally distributed and no population parameters are estimated. These tests are called *nonparametric tests* or *distribution-free tests*.

NULL HYPOTHESIS STATISTICAL TESTING (NHST)

In today's textbooks the process that developed over some 50 years is referred to as *null hypothesis statistical testing* (NHST) and is often treated as a rule-based method of reaching a decision about the effect of the IV on the DV. NHST begins with the two hypotheses about the populations the data are from—the null hypothesis of no difference (H_0) and the alternative hypothesis that the populations are not the same. NHST also begins with a criterion that the researcher chooses for deciding between the two. This criterion is a probability, symbolized α (alpha). The sample data are analyzed with a statistical test that produces a probability figure, p, which is the probability of the data that were observed, if the null hypothesis is true.

If the value of p is equal to or less than the value of α, the null hypothesis is rejected and the alternative hypothesis is accepted. In this case, the test result is statistically significant.

If the value of p is greater than α, the null hypothesis cannot be rejected, so it is retained. This outcome, $p > \alpha$, produces an inconclusive result because both hypotheses remain. Note, however, that although the logic of NHST has not led to a rejection of the null hypothesis, the researcher does have the sample data, which give hints about the populations. In the case of $p > \alpha$, NHST procedures do not permit a strong statement about the relationship of the parameters in the population.

Table 6.3 summarizes the four outcomes that are possible with NHST tests. There are two ways to make a correct decision and two ways to make an error. The probability of a Type I error (α) is chosen by the researcher. The probability of a Type II error is β, which is discussed under the topic of power.

The developers of statistical techniques all emphasized that their procedures were to be aids to decision makers and not rules that lead to a decision. However, subsequent

Table 6.3 Correct decisions and errors that occur with NHST tests

		True situation in the population	
		H_0 true	*H_0 false*
The decision based on sample data	Retain H_0	Correct decision	Type II error (prob. = β)
	Reject H_0	Type I error (prob. = α)	Correct decision

researchers and textbook writers began using NHST logic as an algorithm—a set of rules that lead to a conclusion that is correct with a specified degree of uncertainty. In this rigid view, if $p \leq \alpha$, reject the null hypothesis. If $p > \alpha$, retain the null hypothesis. As mentioned earlier, however, the very basis of the decision rule, sampling distributions and their *p* values, are themselves dependent on the mathematical assumptions that allow sampling distributions to be derived. If the assumptions don't hold, the accuracy of the probability figure is uncertain. It is the case that every statistical test is based on some set of assumptions about the nature of the population data, and some of these sets are more restrictive than others.

ASSUMPTIONS UNDERLYING SAMPLING DISTRIBUTIONS

A sampling distribution is a formula for a curve. Using a sampling distribution, researchers or statisticians determine the probability of obtaining a particular sample (or one more extreme) in a random draw from the population. To derive a sampling distribution, mathematical statisticians begin with a particular population. The characteristics of that population are assumed to be true for the population that the research sample comes from. For example, the sampling distributions that Gosset and Fisher derived for the *t* test and the analysis of variance assumed that the populations were normally distributed. If the populations the samples come from have this characteristic (and others that were true for the populations from which the sampling distribution was derived), then the probability figure the sampling distribution produces is accurate.

Although there are statistical tests to determine the nature of the populations the samples come from, the data available for the test are often limited to the research samples. If these data are ample, these tests work well. If ample data are not available, the tests are of questionable value. However, for some measures that are commonly used by researchers, there is more than enough data to assess the nature of the populations. For example, we know that IQ scores are normally distributed and that reaction time scores and income data are skewed.

The question of what statistical test to use for a particular set of data is not easily answered. A precise solution requires knowing the mathematical assumptions of each test and that the assumptions are tenable for the population being sampled from. Researchers, however, are usually satisfied with a less precise solution that is based on the category of the data and the design of the experiment.

DATA CATEGORIZATION

The recognition that some kinds of data cannot meet certain assumptions led to efforts to categorize data in such a way that the test needed was determined by the category the data fit. Two different category schemes have emerged. Probably the most widely used scheme has three categories: scaled data, rank data, and category data.

Scaled data are quantitative measures of performance that are not dependent on the performance of other subjects. Measures of time, distance, errors, and psychological characteristics such as IQ, anxiety, and gregariousness produce scaled data. *Rank data* (also called *ordinal data*) provide a participant's position among all the participants. Any procedure that results in simple ordering of 1, 2, 3…produces rank data. Situations in which participants order a group of items according to preference, creativity, or revulsion produce rank data. Class standing and world rank are examples of rank data. *Category data* (also called *nominal data*) consist of frequency counts. A variable and its levels are defined and the number of subjects or participants who match the definition is enumerated. A variable such as gender with frequency counts of males and females is an example, as is the number of individuals who check agree, no opinion, or disagree on a Likert scale.

Listing the categories in the order of scaled, rank, and category puts them in the order of more information to less information. Given a choice, researchers prefer to use measures with more information rather than less, but many times the circumstances of their research designs leave them no other practical choice.

The other common way to categorize data is with four scales of measurement. The first scale, *nominal* measurement, corresponds to category data. The second scale, *ordinal* measurement, has the same definition as that for rank data. The *interval* scale of measurement has scaled scores with two characteristics. The first characteristic is that the intervals between equal scores are equal. The second is that the score of zero is arbitrarily defined and does not mean an absence of the thing measured. Common examples of interval scales are the Celsius and Fahrenheit temperature scales. In neither case does zero mean the complete absence of heat and, of course, the two zeros don't indicate the same amount of heat. However, a rise of 10 degrees from 20 to 30 is the same amount of heat as a rise of 10 degrees from 80 to 90 for Celsius and for Fahrenheit scales. The *ratio* scale consists of scaled scores as well, but zero on the ratio scale means the complete absence of the thing measured. Errors (both size and number), reaction time, and physical measures such as distance and force are all measured on a ratio scale.

EXPERIMENTAL DESIGN

Besides the kind of data produced, researchers select a statistical test according to the design of their experiment. In some experiments, each of the scores is *independent* of the others. That is, there is no reason to believe that a score could be predicted by knowing an adjacent score. In other experiments, such a prediction is quite reasonable. Experiments in which one twin is assigned to the experimental group and the other to the control group

Table 6.4 A sampling of NHST tests

Statistical Test	*Data*	*Samples*	*Number of independent variables*	*Levels of independent variable*
Independent *t* test	scaled	independent	1	2
Paired *t* test	scaled	related	1	2
Confidence interval	scaled	independent or related	1	2
One-way ANOVA	scaled	independent	1	2 or more
Factorial ANOVA	scaled	independent or related	2 or more	2 or more
Repeated measures ANOVA	scaled	related	1 or more	2 or more
Analysis of covariance	scaled	independent or related	1 or more	2 or more
Tukey HSD	scaled	independent or related	1	2
Dunnett's test	scaled	independent or related	1	2
Ryan (REGWQ)	scaled	independent or related	1	2
Orthogonal contrasts	scaled	independent or related	1	2 or more
Bonferroni test	scaled	independent or related	1	2 or more
Linear regression	scaled	related	NA	NA
Multiple regression	scaled	related	NA	NA
Mann-Whitney *U* test	rank	independent	1	2
Wilcoxon matched-pairs signed-ranks *T*	rank	related	1	2
Kruskal-Wallis one-way ANOVA	rank	independent	1	2 or more
Wilcoxon-Wilcox multiple comparisons	rank	independent	1	2 or more
Friedman's rank test	rank	related	1	2 or more
Chi square, goodness of fit	category	NA	NA	NA
Chi square, test of independence	category	independent	2	2 or more
Fisher's exact test	category	independent	2	2
Sign test	category	independent	1	2

can be expected to produce two similar scores (although one score may be expected to be larger than the other). Other designs that produce *related* scores are before-and-after studies and matched pair studies in which each participant assigned to the experimental group is matched to a participant in the control group who has similar characteristics.

In addition to the independent/related issue, experimental designs differ in the number of IVs, which can be one or more. Designs also differ in the number of levels of each IV. The number of levels can be two or more.

Researchers often choose a particular NHST test by determining the kind of data, whether the scores are independent or related, and the number and levels of the IV(s). Table 6.4 is a catalog of NHST tests, many of which are covered in undergraduate statistics courses. Those tests that require scaled data have somewhat restrictive assumptions about the populations the samples are from.

ROBUST TESTS

A *robust* statistical test produces accurate probabilities about the population even though the population does not have the population characteristics the sampling distribution is based on. For example, a test that assumes the population is normally distributed is robust if a sample produces a correct *p* value, even if the population that is sampled from is not normally distributed.

A common way to test robustness is to begin with populations of data created with characteristics unlike those the sampling distribution was derived from. A computer program randomly and repeatedly samples from the population, a technique called the Monte Carlo method. With a large number of samples at hand, the proportion of them that produces test values of, say, 0.05 or less is calculated. If that proportion is close to 0.05, the test is robust, which is to say that the test is not sensitive to violations of the assumptions about the population that the test is based on. As might be expected, when robustness is evaluated this way, the results depend on the degree of "unlikeness" of the sampled populations. A common conclusion is that for research-like populations, statistical tests are fairly robust. That is, they are relatively insensitive to violations of the assumptions they are based on. In general, NHST tests are believed to give fairly accurate probabilities, especially when sample sizes are large.

POWER

The best possible result with NHST is to reject H_0 when the null hypothesis is false. In Table 6.3, this fortunate outcome is in the lower right corner. As seen in Table 6.3, the other possible result is a Type II error, the probability of which is β. Thus, when the null hypothesis is false, the probability of reaching a correct decision is 1-β. This probability, 1-β, is the power of a statistical test. In words, *power* is the probability of rejecting a false null hypothesis.

There are two answers to the question, "How much power should a researcher have for an NHST test?" One answer is that the amount of power should reflect the importance of showing that the null hypothesis is false, if indeed it is false. The second answer is to use a conventional rule-of-thumb value, which is 0.80.

Calculating the numerical value of the power of a particular statistical test is called a *power analysis*. It depends on a number of factors, only some of which are under the control of the researcher. For an NHST test that compares two groups, four of the factors that influence power are the following:

1. *Amount of difference between the populations*. The greater the difference between two populations, the greater the chance that this difference will be detected. Although researchers do not know exactly how different populations are, they can estimate the difference using sample means.
2. *Sample size*. The larger the sample, the greater the power of the test to detect the difference in the populations. Sample size is incorporated into the value of the statistical test used to analyze the data.
3. *Sample variability*. The less the variability in the sample, the greater the power of the test. Sample variability, like sample size, is incorporated into the statistical test value.
4. *Alpha* (α). The larger the value of α, the greater the power of the test. To explain, rejecting H_0 is the first step if you are to reject a false null hypothesis. Thus, one way to increase power to 1.00 is to reject H_0 regardless of the difference observed. As seen in Table 6.3, however, rejecting H_0 when the null hypothesis is true is a Type I error. Even so, if a researcher is willing to increase the risk of a Type I error by making α larger, more tests will result in decisions to reject H_0. Of course, for every one of these cases in which the populations are different, the decision will be correct.

The relations among power and these four factors are such that setting values for any four of them determines the value of the fifth. Determining one of the five values is called a power analysis. In practice, researchers use a power analysis to determine the following:

1. *Sample size when planning a study*. Using previous data or conventional rules of thumb, the sample size required for power = .80 (or some other amount) is calculated.
2. *Power of a study in which H_0 was retained*. High power and a retained H_0 lend support to the idea that any difference in the populations must be small. Low power and a retained H_0 lend support to the idea that a Type I error occurred.

OTHER STATISTICAL TECHNIQUES

Most undergraduate statistics courses emphasize statistical techniques that compare the means of two or more treatment groups. However, other techniques are available. Regression techniques permit the prediction of an individual's score based on one other score (simple regression) or on several scores (multiple regression). Structural equation modeling allows researchers to test a theoretical model with data and then use the data to modify and improve the model. Other techniques are used for other purposes. The underlying basis of most statistical techniques is called the General Linear Model. The NHST techniques discussed in this chapter are all special cases of the General Linear Model, which might be considered the capstone of 20th-century mathematical statistics.

Computer-Intensive Methods

The basis of NHST statistics is a sampling distribution. Traditionally, sampling distributions are mathematical formulas that show the distribution of samples drawn from populations with specified characteristics. Computer-intensive methods produce sampling distributions by repeatedly and randomly drawing samples from a pseudo-population that consists of data from the experiment. A particular method called *bootstrapping* illustrates this approach. With bootstrapping, as each score is selected it is replaced (and can be selected again).

To illustrate, suppose a researcher is interested in the mean difference between two treatments. Data are gathered for each treatment, the means calculated, and one is subtracted from the other. If the two samples are from the same population (the null hypothesis), the expected difference between the two sample means is zero. Of course, two actual means will probably differ. To generate a sampling distribution that shows the entire range of differences, all the sample data are used to create a pseudo-population. A computer program randomly draws many pairs of samples, calculates the two means, and does the subtraction. The result is a sampling distribution of mean differences. Finally, the difference between the two treatment means obtained in the experiment can be compared to the mean differences in the sampling distribution. If the difference obtained in the study is not likely according to the sampling distribution, reject H_0 and conclude that the two treatments produce data that are significantly different. Thus, computer-intensive methods are also based on NHST logic.

Meta-Analysis

Meta-analysis is a statistical technique that addresses the problem of how to draw an overall conclusion after researchers have completed several studies on the same specific topic. It is not unusual for a series of studies to produce mixed results; some reject the null hypothesis and some retain it. Meta-analysis is a statistical technique pioneered by Glass (1976) that combines the results of several studies on the same topic into an overall effect size index. This averaged effect size index (often an averaged *d* value, $\bar{d}$) is reported with a confidence interval about it. A $\bar{d}$ value is interpreted in the same way as a *d* value from one study. The confidence interval about $\bar{d}$ is interpreted in the same way as a confidence interval about a mean.

The NHST Controversy

In the 1990s a controversy erupted over the legitimacy of NHST. The logic of NHST was attacked, and researchers who misinterpreted and misused NHST were criticized. One of the central complaints about the logic of NHST tests was that the best they could provide was a conclusion that could be arrived at logically without any data analysis. The case can be made that for any two empirical populations, the null hypothesis is always false if you carry enough decimal places. If this is true, the best that NHST can do is to confirm a conclusion that could have been reached on logical grounds—that the populations are different.

A second complaint about the NHST procedure is that it always starts at square one, ignoring any previous work on the variables being investigated. As a result, a sequence of experiments on the same variables, analyzed with NHST techniques, does not lead to more precise estimates of parameters, but only to more confidence about the ordinal positions of the populations.

A third problem with the logic of NHST is that it focuses on minimizing the probability of a Type I error, the error of rejecting a true null hypothesis. Of course, if two populations are different (as noted in the first complaint), Type I errors cannot occur. In addition, focusing on Type I errors leaves evaluation of Type II errors as an afterthought.

The most prevalent misunderstanding of NHST is of the *p* in $p < 0.05$. The *p* is the probability of the test statistic when the null hypothesis is true. It is not a probability about the null hypothesis. It is not the probability of a Type I error, the probability that the data are due to chance, or the probability of making a wrong decision. In NHST, *p* is a conditional probability, so *p* is accurate only when the condition is met, and the condition is that the null hypothesis is true.

Another complaint, alluded to earlier, is about the practice of using the 0.05 alpha level as a litmus test for a set of research data. The practice has been that if an NHST test does not permit the rejection of H_0, the study is not published. This practice can deprive science of the data from well-designed, well-run experiments.

As a result of the complaints against NHST, the American Psychological Association assembled a task force to evaluate and issue a report. Their initial report (http://www.apa.org/science/bsaweb-tfsi.html) recommended that researchers provide more extensive descriptions of the data than just NHST results. They recommended increased reporting of descriptive statistics, such as means, standard deviations, sample sizes, and outliers, and a greater use of graphs, effect size indexes, and confidence intervals. In their view, NHST techniques are just one element that researchers use to arrive at a comprehensive explanation of the results of a study and should not be relied on exclusively.

THE FUTURE

As the reader has no doubt detected, the field of statistics has been a changing, dynamic field over the past one hundred years. The future appears to be more of the same. As an example, Killeen (2005) described a new statistic, p_{rep}, which has nothing to do with null hypothesis statistical testing. The statistic, p_{rep}, is the probability of replicating the results that were observed. In particular, p_{rep} is the probability of finding a difference with the same sign when the control group mean is subtracted from the experimental group mean. Thus, if the original experiment resulted in $\bar{X}_e = 15$ and $\bar{X}_c = 10$, then p_{rep} is the probability of $\bar{X}_e > \bar{X}_c$ if the experiment were conducted again.

This new statistic, p_{rep}, is one response to the complaints about NHST logic described earlier. It has a number of advantages over the traditional *p* value from a NHST analysis, including the fact that because there is no null hypothesis, there are no Type I and Type II errors to worry about. It is a continuously graduated measure that does not require a dichotomous reject/not reject decision. Finally, the calculation of p_{rep} permits the incorporation of previous research results so that the value of p_{rep} reflects more than just the present experiment. The statistic p_{rep} appears to be catching on, showing up regularly in *Psychological Science*, a leading journal.

As the 21st century began, the prevailing statistical analysis was NHST, a product of 20th-century statisticians. As for the future, new statistics are always in the works. NHST may not remain the dominant method for analyzing data; other statistics may supplant it, or it may continue as one of the techniques that researchers use to analyze their data.

REFERENCES AND FURTHER READINGS

Fisher, R. A. (1925). *Statistical methods for research workers*. London: Oliver and Boyd.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3–8.

Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, *16*, 345–353.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one or two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50–60.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient test of statistical hypotheses.

Philosophical Transactions of the Royal Society, A 231, 289–337.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century.* New York: W. H. Freeman.

Student. (1908). The probable error of a mean. *Biometrika*, *6*, 1–26.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics, 1*, 80–83.

7

VALIDITY

KENNETH BARRON, WITH ALLISON R. BROWN, THERESA E. EGAN,
CHRISTOPHER R. GESUALDI, AND KIMBERLY A. MARCHUK

James Madison University

If we had to identify one concept that we hope all psychology students would master after taking their coursework in research methodology, it would be *validity*. Put simply, validity represents accuracy, credibility, and soundness. It is derived from the Latin word *validus*, which means strength. When applied to psychology, validity provides a vital tool to judge the quality of our scientific evidence about psychological phenomena.

However, despite the important role that validity plays in psychology, it can be easy for new students to get lost in the different terms and jargon associated with it. Therefore, the purpose of our chapter is twofold. We begin with a class exercise that we use to introduce validity and its role in the scientific process. Then, we review the common validity terms that appear in psychology and advance an overarching framework of validity to simplify and organize the different terms and jargon associated with validity.

AN INTRODUCTORY EXERCISE ON VALIDITY

As an initial exercise to introduce validity, we ask students to read the following article from the University of Wisconsin's school newspaper, *The Badger Herald*. So you can experience this activity firsthand, a shortened version of the article appears below. Take a few moments to read the article, and we'll pose to you the same questions we ask our students.

A Recent Survey Conducted by the Princeton Review Named UW, Madison, the Number 2 Party School in the Country*

By Kira Winter, Feature Editor

Welcome to UW, Madison—the No. 2 party school in the nation. UW was recently bestowed this high honor by the Princeton Review, finishing only behind West Virginia University, Morgantown.

UW lived up to a "study hard, party hard" reputation in the minds of 56,000 surveyed students.

Although UW Junior Lindsay Moore does not disagree that she attends a school where students like to have a good time, she is leery about the reliability of the findings. "There are surveys for just about everything," Moore said. "I think if someone came here on a football weekend they would see a very different campus than if they came in the dead of winter."

Moore's mother, Gail, thinks students at every college or university party, but UW's heritage singles it out. "Drinking, especially beer drinking, is definitely part of the Wisconsin tradition and people play into that," she said.

How do incoming freshmen feel about maintaining the school's wild reputation? "I think it's great that UW was voted the number two party school and I am definitely excited to contribute," said freshmen Andrea Santiago.

*Winter, Kara. (1997, August 29). A recent survey conducted by the Princeton Review named UW-Madison the number 2 party school in the country. *Badger Herald*, University of Wisconsin, Madison. Reprinted by permission of the *Badger Herald.*

> And what about those concerned parents, leaving their innocent children in the clutches of such wide-eyed hedonism? Dan and Joanne Schwartz were slightly apprehensive about leaving their son, an entering freshman, after they heard the survey on the news, but think UW students maintain a good balance of studying and partying. “To be honest, we had a hard time sleeping the night we heard that it was the number two party school on the national news,” said Schwartz’s father. “But we know it’s a great school academically, and fun, too, so we really aren’t too worried.”
>
> Many school administrators are taking the new honor with a grain of salt. “Of course the issue of alcohol use is something we are definitely concerned with,” said Roger Howard, Associate Dean of Students. “However, I don’t think excessive alcohol use and being voted the number two party school are really related. It has more to do with the enthusiasm and energy of the students.”
>
> Top 10 Party Schools: 1. West Virginia University, Morgantown, 2. University of Wisconsin, Madison, 3. SUNY, Albany, 4. University of Colorado, Boulder, 5. Trinity College, 6. Florida State University, 7. Emory University, 8. University of Kansas, 9. University of Vermont, 10. Louisiana State University.

Now, consider the following questions: Is the statement that UW is the No. 2 party school in the nation *valid*? If yes, which elements give the article credibility to support this conclusion? If no, which elements have you questioning the conclusion that UW is the No. 2 party school?

When we conduct this exercise, students identify several things from the article that appear to support the validity that UW is the No. 2 party school. For example, they note that the findings are based on a large sample of 56,000 students. They also note that the study was conducted by the Princeton Review, which students regard as a credible source because it is well known for publishing college guidebooks and SAT prep books. However, they also identify several things that call into question the validity of this finding. For example, they note that one of the students quoted in the article stated that a different impression of the “party” atmosphere of the Wisconsin campus could be formed depending on when you visited the campus. Students also indicate that the label of “party school” may carry different connotations for different people, thereby influencing results. For example, although the UW students and parents interviewed in the article appeared to connect a “party school” to the consumption of alcohol, the Associate Dean of Students noted that UW students are enthusiastic and energetic. We will return to this article as a case study throughout the chapter.

WAYS OF KNOWING

Interestingly, before critically discussing the article and questioning the validity of its findings as a class, many students admit that they accepted the conclusion that UW was the No. 2 party school. Charles Peirce, a 19th-century American philosopher of science, outlined several different strategies that individuals employ as they come to accept something as true. (These have been popularized as the “ways of knowing.”) One way is through *tenacity*, which is the acceptance of a belief based on the idea that we have always known it to be that way. For example, when reflecting on our newspaper article about UW being the No. 2 party school, one of the parents interviewed noted, “Drinking, especially beer drinking, is definitely part of the Wisconsin tradition and people play into that.” When beliefs are based on tenacity, they are difficult to change, but when individuals are objective and open to new data, these beliefs may be disconfirmed.

A second way of knowing is through *authority*, which means accepting a belief because authority figures (such as parents, teachers, or experts) tell us to. In the case of UW being No. 2, the article was based on information reported by the Princeton Review, which could be deemed an authority about university life. However, authority figures may be wrong or have hidden agendas.

A third way of knowing is through *rationalism*, which is the acceptance of a belief through the use of logic and reasoning. Again, one may logically deduce a lot of partying occurs at UW through deductive reasoning—for example, people party at bars. Madison, Wisconsin, has lots of bars. Therefore, Madison has lots of parties. Rationalism has appeal because we are now using our own independent thinking and logic to form an opinion. However, what may make sense in theory may be far from true in reality.

A fourth way of knowing is through *empiricism*, which is the acceptance of a belief based on direct personal observation. If you haven’t been to UW, a visit could reveal many highly visible places in which you could observe students engaged in partying. For example, at the time that this article was published, UW’s Student Union still served alcohol to students, faculty, and staff. However, direct experience also can be biased and misleading. Note how one student quoted in the article indicated you might get a far different impression if you visited on a football weekend versus another time of year, especially in the dead of winter.

However, Peirce was quick to point out how each of the previously mentioned ways of knowing can be misleading. Each can be subjective and influenced by idiosyncratic or cultural biases, they can be marked by an uncritical and accepting attitude to gaining knowledge, and, in the case of the first three ways of knowing, no systematic search for verifiable truth is made by the individual. Thus, knowledge obtained from these methods may or may not be true, but you won’t have any way of telling the difference. His solution was to use *science* as a fifth and final way of knowing,

When adopting a scientific approach, we combine rationalism and empiricism. Rationalism (using logic and reasoning) is related to theory-building and generating hypotheses. Science often begins by devising a “theory” to explain a particular phenomenon. Empiricism (using personal observation) is related to testing theories and hypotheses. As a way of knowing, a scientific approach systematically evaluates whether the prediction of a given theory will hold true in the observable world. "As one leading psychological researcher put it, the scientific method…helps us sort out what we know about human nature from what we only *think* we know" (Rosnow & Rosenthal, 1996, p. 6).

JUST BECAUSE YOU ENGAGE IN SCIENCE DOESN'T MEAN IT'S GOOD SCIENCE

However, just engaging in the scientific method as a way of knowing doesn't guarantee good science! We still need a framework to determine if a study's conclusions are valid and to appreciate that not all science is "created equally." Scientific studies may vary tremendously in the quality of information and strength of evidence that they provide. Thus, another defining feature of science is to adopt a skeptical and critical attitude to judge the quality of evidence that we collect. Concepts of validity provide that framework, and if you take the time to master those concepts in your psychological training, you will place yourself in an excellent position to strengthen your ability to separate good science from bad science. It also will provide you with the ability to determine what science is lacking and what should be done to improve on what we currently know.

Mastering validity, however, requires learning a new vocabulary. Validity is an area of science in which a wide range of terms is commonly used. For example, we conducted a review of 15 popular undergraduate psychology research methods textbooks to evaluate how many terms are typically reported. Individual textbooks described from 4 to 12 different types of validity, and 15 unique types of validity were discussed across all textbooks. Making matters worse, you may encounter researchers who use *different* terms to describe the *same* validity concepts. Or you may encounter researchers who use the *same* terms even though they are talking about *different* validity concepts.

A common critique of many psychological fields concerns the proliferation of terms used to describe various concepts, especially as work in that field grows over time. Thus, an important call for any field is to balance the number of terms and concepts used to capture the complexity of the phenomena with striving for parsimony and the simplest explanation (which in this case would be the fewest validity terms necessary). In addition, if we can avoid redundancy and overlap in our language, we can develop a simple, common language that can be quickly learned, providing a universal dialogue in which to converse.

It is with this goal that we would like to simplify the complexity found in validity terminology. Following Shadish, Cook, and Campbell's (2002) guidelines, we recommend a validity framework in which the overarching themes of validity can first be grouped into four major areas. Then, we will share our review of the research literature on validity for other terms that have appeared that can fall under the umbrella of one of these four overarching areas. With our students, we affectionately refer to this organizational framework as the Big 4 model of validity (modeled after the Big 5 model in personality). These areas entail *construct validity*, *external validity*, *internal validity*, and *statistical conclusion validity* (see Cook & Campbell, 1979; Shadish et al., 2002).

CONSTRUCT VALIDITY

Construct validity involves making a systematic evaluation of the accuracy and strength of the constructs used in the study. A *construct* is simply a synonym for the research variable being investigated. Thus, construct validity entails asking the following question about a study: "Does the study accurately represent and evaluate the constructs (aka variables) under investigation?"

To diagnose if a study is weak or strong in construct validity, you first need to determine each of the constructs that the researcher is interested in studying. Then, for each of these constructs, you want to make an independent judgment on whether you think the researcher has accurately represented that construct. For example, in our opening case study, the Princeton Review was interested in finding out how universities differed with respect to student partying. In this case, there is only one main construct being studied: partying behavior. However, an essential component to any scientific research study is converting a research question about a theoretical idea (like partying) into a measurable way to assess and evaluate the construct. When we critically evaluate the construct validity of our opening article, we need to consider how the researchers defined and measured partying.

Turning back to the article about UW, can you determine what conceptual definition the Princeton Review used for partying, and what actual measure the Princeton Review used to determine it? Unfortunately, when the media reports research, detailed descriptions of the actual measures used are rarely included. Instead, we are left to brainstorm the possible ways in which partying behavior could have been measured. For example, does partying mean the number of social events a student attends while in college? Does partying mean the number of actual hours a student spends partying? Or does partying mean engaging in a certain kind of activity, like drinking alcohol, which may be in fact what the Princeton Review wanted to capture? Depending on which definition of partying is used, we may find a rather different answer regarding which university students' engage in the most partying. So, we might quickly critique that "party school" is an ambiguous term that should be replaced with a more descriptive term such as "alcohol use" or "social drinking," if that was the goal of the Princeton Review's study of college campuses.

Unlike popular media sources, academic reporting of research involves a much more detailed and in-depth explanation of what a researcher did to make clearer evaluations. Researchers also have many additional types

Figure 7.1 Specific terms and concepts associated under the umbrella of construct validity.

of validity and types of research tools that help them argue that they truly have accurately represented the psychological constructs that they are interested in investigating. Indeed, there is an entire psychological field focused on developing tools for the accurate measurement of psychological phenomena (Benson, 1998; Bryant, 2000; Cronbach, 1990; Messick, 1995; Shepard, 1993). To help simplify the complexity of what you may encounter and read, we would like to place these additional terms under the broader umbrella of construct validity (see Figure 7.1). These concepts are subtypes of construct validity. As researchers, we look for this type of information in academic journals when we critique the construct validity of a study. We want to see the researcher convince us that the measures that have been adopted will provide a strong and accurate measure of the constructs that the researcher is investigating.

There is no doubt that construct validity is the most complex of our Big 4 validities to describe because of its many different subtypes. So, let's take another example of a measured variable important for getting into many colleges, the SAT, which is a construct that is argued to reflect students' college preparedness and potential academic aptitude. This example will help you appreciate how researchers evaluate additional information about a measure in order to argue that it is high or low in construct validity.

One source of evidence is *content validity*, which reflects the extent to which a measure appropriately covers the full breadth and range of the construct under investigation. If you use a scholastic aptitude test to measure what you've learned in school, then what content should be on that test if you want to argue the measure is high in content validity? Should it just have math, verbal, and writing components like the SAT? Or should it reflect the broader range of academic subjects that you take in school (including content from the natural sciences, social sciences, arts and humanities, and foreign languages)? Right away we may critique the SAT for being a narrowly focused test. Thus, it lacks content validity, which has stimulated the development of other, broader-based tests such as the ACT. The narrow focus of the SAT is one reason why colleges and universities often ask students to take other achievement tests.

Another source of evidence is *convergent validity*, which is a technique that involves comparing your measure to other known measures to which it should be theoretically related. For example, the SAT could be compared to other known college entrance assessments such as the ACT, and we would predict that students' scores should be positively related if in fact both reflect the same underlying theoretical constructs. However, equally important can be demonstrating *discriminate validity* by showing how your measure is distinct and unique from other measures. For example, if we wanted to demonstrate that the SAT reflects a specific aptitude test rather than a more general IQ test of overall intelligence (another popular notion of what the SAT evaluates), we could compare students' scores on the SAT scores to IQ scores. If we are correct, SAT and IQ scores should be less related to each other than SAT and ACT scores. By showing how our measure both relates to and doesn't relate to other known measures, we are in a better position to validate what our construct represents.

Another source of evidence is *criterion-related validity*, which reflects the extent to which a measure accurately relates to other criteria or outcomes to which you would theoretically predict it would be related. These other criteria may be about things in participants' past, present, or future. Researchers will then compare scores on their measure with these criteria to see if they align with their predictions. For example, SATs could be compared to students' high school GPA to look at criteria in the past (this type of criterion validity is known as *retrospective validity* because you are going back in time retrospectively to acquire information that should be related to your construct). We might predict that if you do well on the SAT, you also probably did well in high school. Alternatively, the SAT score could be compared to students' future college GPA because you might predict that students with higher academic aptitude coming into college would perform better in college (this type of criterion validity is known as *predictive validity* because you are going forward in time). The SAT score also could be obtained at the same time with other measures that you believe are associated to the SATs, like the ACT (this type of criterion validity is known as *concurrent validity* because measures are all collected at the same time).

Face validity evaluates how accurate a measure seems on its appearance to measure the construct of interest (its face value). So, after reviewing test items on the SAT, you would likely conclude that it appears to assess math and verbal aptitude. If the items assessed other content but the researchers still argued that it only assessed math and verbal aptitude, you might question the face validity of the test. Finally, *known-groups validity* is a technique in which a researcher gives a measure to known groups that should differ on the measure. For example, scores on the SAT should be higher for a group of students who take college prep work in high school, and scores should be lower for a group of students who dropped out of high school. If the groups differ in the predicted way, then the measure will appear to be more accurate. Although there are many different subtypes of construct validity, the take-home message is appreciating that these additional validity types are research tools to aid us in making an overall judgment about construct validity.

The accurate measurement of constructs also is tied to a related concept known as *reliability*. Although reliability often appears as its own topic in research method textbooks, we argue that it, too, falls under the broader umbrella of construct validity. Whereas validity is associated with overall accuracy and strength, reliability is associated with consistency. Researchers wish to make sure their measures consistently assess the behavior in which they are interested to assure that their measures are valid. If their measures are unreliable, then the construct validity of their measure is quickly jeopardized. Just as there are many subtypes of validity, there are many subtypes of reliability that researchers report to help support that they have collected a reliable measurement. The types of reliabilities that are reported depend on the nature of the measure that the researcher uses. The three most common types of reliability are *test-retest reliability*, *internal reliability*, and *interobserver reliability*. For example, if we argue that the SAT reflects a stable measure of scholastic aptitude in math and verbal domains, then we would expect that we would have similar scores if a person took the test today and then again next week. This type of reliability is known as *test-retest reliability*. When measures comprise multiple items that form an overall index and score of that construct, another index of reliability assesses how consistently all of the items of a measure assess the same construct. This type of reliability is known as *internal reliability* or *internal consistency.* For example, the SAT verbal score is made up of multiple items, and thus we would hope to see that the test items have high internal reliability. A test with low internal reliability suggests that participants are not responding consistently to the items and that the set of questions fails to assess the same construct. Or, if a measure is collected by having observers watch and record a behavior, another helpful type of reliability is *interobserver reliability*. Rather than having one person observe and score a behavior, researchers benefit by having multiple people observing and then assessing inter-rater reliability by comparing how similar the ratings are from different observers.

So far we have covered one major type of construct found in research, variables that are *measured*. However, researchers also are interested in a second type of construct: variables that are *manipulated*. For example, we asked you to consider the construct validity of SAT tests, and whether the SAT reflects a valid measure that should be used in college admissions. This example involves only a *measured* variable. But now consider a study in which a researcher is interested in evaluating a manipulation that could improve students' SAT scores. For instance, a researcher may be interested in formally testing whether students who receive SAT coaching before taking the SAT perform better than students who do not receive coaching. One form of SAT coaching would be buying a study guide for the SAT before taking the test. Indeed, another major facet of the Princeton Review (in addition to writing college review guides) is to provide prep guides for college entrance tests. Now, we have an example of a study with a *manipulated* variable (receiving or not receiving SAT coaching) and a *measured* variable (SAT score). Just as we are interested in evaluating the construct validity of measured variables, we are equally interested in evaluating the construct validity of manipulated variables.

One effective tool to determine if researchers have successfully manipulated a particular variable is to conduct a *manipulation check* to assess if the treatment functioned as the researchers intended. For example, if coaching entailed studying a SAT prep guide, then one possible manipulation check question could be asking whether study participants used prep guides before the exam. If we found that students in the treatment group used their prep guides while students in the no-treatment group refrained from using any prep guides, we would have construct validity evidence that our manipulation was accurately followed by participants. Alternatively, if we found that students in both groups used prep guides or received other types of coaching, then we would have evidence that our manipulation was contaminated and failed to represent the manipulated construct we intended.

In sum, construct validity entails separately evaluating whether researchers have accurately measured and/or manipulated each of their variables under investigation. If we think one or more variables are suspect in accurately representing what the researchers want represented, then we would critique the study on its overall construct validity. We have covered a lot of material, and once again it could be easy to get lost in terms and jargon. However, to help reinforce your ability to literally see the "bigger picture," look again at Figure 7.1 to see how construct validity provides an overarching umbrella to organize and synthesize a wide range of associated concepts.

EXTERNAL VALIDITY

External validity involves making a systematic evaluation of the accuracy and strength of the ability to generalize the results beyond the current study. When assessing external validity, we typically ask if the study's results can be generalized and applied to other people, settings, or times.

Once again, many related validity terms and subtypes of external validity have been proposed that we would like to organize under the overarching umbrella of external validity (e.g., Bracht & Glass, 1968). Specifically, *population validity* entails the extent to which findings can be generalized beyond the participants of a study to the population at large. *Ecological validity* entails the extent to which findings can be generalized across different settings, particularly more naturalistic, real-world settings. *Temporal validity* entails the extent to which the findings of a particular study can be generalized across time. For example, would the results of a study conducted in 1950 still apply and generalize to today if the study was replicated? Rather than talking about external validity in broad terms, these additional concepts have appeared to help researchers highlight a specific facet of external validity focused on generalizing across people, settings, *or* time.

In addition, new external validity terms have been coined to capture unique generalizability issues in particular fields. For example, in industrial/organizational psychology, the terms *intraorganizational validity* and *interorganizational validity* (Goldstein, 1991) emphasize whether the practices found to work in one organization/company can generalize to other parts of the same organization/company (intraorganizational validity), or if findings can generalize to other organizations/companies (interorganizational validity). This should help you appreciate how validity tools evolve and are added to the research literature to help hone our ability to stress what is and isn't strong about a particular research study.

Different psychological fields have also called into question the external validity of their findings. For example, Sears (1986) wrote an influential article, titled "College Sophomores in the Laboratory: Influence of a Narrow Data Base on Social Psychology's View of Human Nature," that questioned the external validity and generalizability of social psychological research. In particular, Sears conducted a review of research published in the top social psychology journals in 1980 and 1985 and coded each study for the type of participants and the type of settings used. He documented an alarming trend: Researchers were overwhelmingly conducting studies on one particular type of participants (with over 80 percent involving college students) in one particular type of setting (with over 70 percent in an artificial, laboratory environment). This trend was in stark contrast to research conducted a quarter of a century earlier in social psychology that tested more diverse types of participants in more naturalistic, real-world settings. Sears provided compelling evidence to question the external validity of many of social psychology's findings and suggest how the unique characteristics of oversampling college students in laboratory settings may bias and limit our understanding of psychological phenomena. As a result, he challenged the field to return to conducting research that would allow stronger claims of external validity to be made.

If we return to our opening case study of UW being the No. 2 party school, how would you critique the study on its external validity? Remember, our students often report being impressed that the Princeton Review surveyed 56,000 students. However, an important conclusion of taking coursework in statistics and research methods is learning that bigger samples aren't necessarily better. One essential key to conducting research and being able to generalize back to a larger population is to ensure that we *randomly sample* participants from the population that we are interested in studying. Random sampling ensures that each participant in the targeted population has an equal chance of being included in the study. If we randomly sample, we can be more confident that we have captured a representative sample of the population and that the results based on our sample can be generalized to the population as a whole. In contrast, many researchers engage in *convenience sampling*, where the sample is simply made up of participants who are easily accessible and willing to become involved with the study. Unfortunately, convenience sampling can lead to a host of systematic biases of who is studied. Because a convenience sample is unlikely to reflect accurately all the characteristics of a given population, we are in a weak position to generalize back to the larger population.

As it turns out, the Princeton Review only used convenience sampling. They simply went to different college campuses and set up survey booths in high-traffic areas on campus (e.g., outside a student union) and asked students who were willing to fill out surveys about their college. Would students who took the time to stop to fill out the surveys accurately represent what all students on that campus thought about their college? Because the answer is clearly "no," the Princeton Review study lacks population validity. Then, regarding temporal validity, recall the student quoted in the party school article who noted that the timing of a visit (a football weekend versus the dead of a Wisconsin winter) would likely produce very different impressions of the campus. And finally, across the years, UW's No. 2 ranking has fluctuated. The Princeton Review reported UW was the No. 2 party school in 1997 and 1998, fell to No. 20 in 1999, but was back to No. 2 in 2003. Thus, to help ensure external validity, researchers need to consider carefully how to sample their participants and to consider testing a wide array of situations across time if they want to argue that their study is high in external validity. Figure 7.2 shows a summary.

Figure 7.2 Specific terms and concepts associated under the umbrella of external validity.

INTERNAL VALIDITY

Internal validity is an evaluation of a special type of "internal relationship" that can occur between variables in a study. Specifically, internal validity evaluates whether a cause-and-effect relationship between variables can be determined. Cause-and-effect conclusions hold a special place in psychology because we are ultimately interested in trying to explain the underlying cause of behavior. If a researcher does X, will it cause Y to occur? For example, if a college campus is afraid of a growing alcohol problem, can a treatment be administered on campus with the goal of decreasing the amount of drinking that occurs?

The 19th-century philosopher of science, John Stuart Mill, argued that three criteria are needed if we are to deduce that two variables are causally linked. First, if one event causes another, then the cause must precede the effect in time. This idea is called *temporal precedence*. Second, if one event causes the other, then the two events must covary together—if one changes, the other must change too. This idea is called *covariation of events*. For example, if X occurs, then Y occurs. Similarly, if X doesn't occur, then Y doesn't occur either. Third, if one event causes the other, you want to be certain that the hypothesized cause is the only cause operating in the situation to create the effect. This idea is called *elimination of plausible alternative causes*. If another variable could have created the same effect, then you cannot make a strong causal inference.

Researchers attempt to meet these criteria through experimental research designs. An experiment allows researchers to manipulate when X occurs and when Y is measured (establishing temporal precedence). An experiment has at least two groups being compared to each other, typically where one group gets the experimental treatment and another serves as a nontreatment comparison group (i.e., control group). Comparing an experimental group and a nontreatment group allows a researcher to see what happens to Y when X is manipulated (establishing covariation of events). Finally, in an experiment, researchers strive to control extraneous variables that also may cause changes in Y (ruling out plausible alternative causes). To gain this level of control, researchers often conduct their research in controlled laboratory settings rather than in more naturalistic field settings, which are far more difficult to control in order to rule out other extraneous variables. In fact, this methodological decision quickly highlights a trade-off that frequently occurs between validities: The steps that are taken to improve internal validity often limit and reduce a study's external validity.

If we return to our opening case study, how would you critique the study on its internal validity? First, internal validity is about establishing a cause-and-effect relationship between two or more variables, therefore a research study must have multiple variables under investigation and have a proposed research question that X is the cause of Y. Second, the only type of study that can establish clear cause-and-effect relationships is an experiment. If nonexperimental methods are adopted (e.g., descriptive or correlational methods), then the study is by default weak in internal validity. Finally, if an experiment is conducted, all three of Mill's criteria for establishing cause-and-effect claims need to be met.

A review of the case study quickly reveals that we are unable to meet these required conditions, and thus it is low in internal validity. For example, the case study involves only one measured variable, partying, therefore there is no proposed cause-and-effect question to be addressed. In addition, the methodology of our case study is nonexperimental. The Princeton Review simply surveyed students about their partying behavior. However, the high rates of drinking on college campuses have led many universities to engage in more experimental designs to evaluate and test the effectiveness of different types of treatments to reduce alcohol consumption and to teach students to become socially responsible drinkers (e.g., Wechsler et al., 2003). In this case, we would have multiple variables under investigation in the form of a cause-and-effect research question to test if the campus' alcohol intervention (i.e., manipulated X variable) is the cause of a reduction in alcohol consumption (i.e., measured Y variable). Figure 7.3 gives a summary picture of our overarching internal validity umbrella and related concepts to determine internal validity.

STATISTICAL CONCLUSION VALIDITY

Our final class of validity is *statistical conclusion validity*, which evaluates the accuracy and strength of the data

Figure 7.3 Specific terms and concepts associated under the umbrella of internal validity.

analyses and statistical conclusions that are drawn from a study. If you were ever wondering why psychology majors take courses on statistics, it is so we can learn how to analyze variables appropriately and correctly and draw conclusions from our research studies. In particular, we have a number of important tools to judge if a study is high in statistical conclusion validity (see Wilkinson & Task Force on Statistical Inference, 1999).

First, we can judge whether the study used the appropriate statistics to summarize and analyze the data. A researcher has many choices on which statistics to use and report. However, for a particular statistic to be appropriate it needs to meet the *statistical assumptions* that are necessary to use it.

Second, after the appropriate tests are conducted, we need to determine whether the researcher has made a *correct statistical decision* or committed a *statistical error*. Statistics allow us to make informed decisions about data. However, statistics are based on probabilities, and thus there is the possibility of arriving at a false conclusion. In particular, training in statistics will teach us how to avoid *Type 1 errors* (detecting a statistically significant effect in our data when one does not exist) and *Type 2 errors* (failing to detect a statistically significant effect when there truly is one).

Third, related to committing statistical decision errors (specifically, Type 2 errors) is another important aspect of statistical conclusion validity that involves having adequate power to test research ideas. *Statistical power* represents the degree to which a study is able to evaluate a statistical effect and to find it reliably when it truly exists. Unfortunately, many research studies can be grossly underpowered, and thus are prone to Type 2 errors.

Finally, an important component of statistical conclusion validity is moving beyond simple statistical significance tests and reporting the overall *effect size* of a study. Effect size provides a direct measure to determine if the effect of a study is small, moderate, or large in magnitude, and reflects the practical or meaningful significance of a study.

Returning one last time to our opening case study, you may find that critiquing the study on its statistical conclusion validity may be the most difficult validity of the four to judge. First, having some background with statistics is necessary before you can begin to critique the statistics of a study. Second, even with this background, popular media reports of research studies rarely provide detailed accounts of the statistical analyses used by the researchers who conducted the study. Therefore, offering a judgment about the strength of the statistical conclusion validity is extremely difficult. Instead, it is easier to offer a critique of what we would need to see before being able to make any definitive conclusions. For example, we would need to know more about the data collected by the Princeton Review to determine what assumptions were met and what statistical analyses were appropriate and inappropriate to use. Figure 7.4 provides a summary picture of our overarching umbrella for statistical conclusion validity and related validity terms and concepts.

SUMMARY

Although there are numerous terms and concepts used to describe validity, we can simplify the complexity of validity by organizing terms and concepts into four main types: *construct validity*, *external validity*, *internal validity*, and *statistical conclusion validity*. Mastering these Big 4 types of validity provides a comprehensive framework to judge the strengths and weaknesses of scientific research.

Determining the overall validity of research is simply not a *black-and-white decision*. Instead, we have four types of validity to reinforce the notion that we have many different ways to determine the strength of a research study. In fact, rarely will you ever encounter a study that ranks strong in all four validities. Remember, the methodological decisions made to increase one type of validity often harm another type of validity (e.g., steps taken to increase internal validity often harm external validity), and similar tradeoffs can occur among all four of our major classes of validity (Mitchell & Jolley, 2006).

In addition, many studies are conducted that do not address all four validities. Recall that only an experimental research methodology is able to establish internal validity and cause-and-effect conclusions. Thus, when nonexperimental research methodologies are adopted, a study by default will be weak in internal validity. Similarly, many researchers have aptly noted that external validity can be a premature goal of research (Mook, 1983). Researchers may

Figure 7.4 Specific validity terms and concepts associated under the umbrella of statistical conclusion validity.

first want to study a phenomenon in isolated populations or isolated situations before trying to generalize their results to the population from which they drew their sample. Therefore, to determine the overall merit of a study, we initially need to recognize what the goals of the study being conducted are and what validities need to be established to meet those goals. The next step would be to weigh the strengths that make the study valid against the weaknesses that threaten the study's validity.

Because individual studies are likely to be weak on one or more types of validity, researchers rarely conduct and report single studies. Instead, they often conduct multiple studies to answer the same research question using different research methodologies. A validity weakness threatening the conclusions of one study can be tested in a new study using a different methodology to address (or remove) that potential threat to validity. A common practice is for researchers to engage in systematic research in which a series of studies are conducted and published over time. Alternatively, researchers will report multiple studies in a single professional article to demonstrate how different threats to validity were tested and ruled out across different studies of the article.

A common area of confusion that can occur when learning about validity is determining which type of validity is being threatened. We have found that students particularly have difficulty when trying to distinguish construct validity problems from internal validity problems. Our first tip is to highlight the fundamental difference between these two ways to critique a research study. A critique of construct validity involves thinking about each variable of a study separately, and then making a judgment on whether each variable was measured or manipulated validly. If you have concerns that a particular measure or manipulation was biased in representing what it was supposed to represent, then your concern involves a critique about construct validity. In contrast, internal validity involves thinking about the relationships among variables of the study, and then making a judgment that one or more of them can be clearly linked to causing other variables in the study. Our second tip is to highlight that each of our validity concepts provides a guiding framework on how to evaluate a study. However, once again, diagnosing validity isn't always a black-and-white decision. Instead, a particular issue about a study may influence multiple validities, blurring some of the clear boundaries that fit nicely under one, and only one, validity label.

Don't be surprised when learning about validity if only *some* of our Big 4 validities are discussed. Our preference is to present validity and the Big 4 types of validity in an integrated fashion to give students a coherent picture to compare and contrast key validity concepts. However, when reviewing undergraduate psychology research methods textbooks, only half of the textbooks presented validity concepts together in an integrated fashion. Just as common was seeing different validity concepts presented in separate chapters. Moreover, although 93 percent of the textbooks discussed construct, external, and internal validity, only 40 percent discussed all four validity concepts. Also, don't be surprised if certain readings about validity focus exclusively on a particular facet of validity such as construct validity (e.g., Benson, 1998; Messick, 1995; Shepard, 1993).

Research methodology and terminology evolve over time, so you may encounter new validity terms that have been developed or validity terms specific only to certain disciplines within psychology. Try to place the new term under one of our Big 4 overarching umbrella terms to understand its purpose and how we can use it to judge a study.

We encourage you to continue to be a student of validity. We have a number of additional readings in our references that we recommend. In particular, we would like to highlight Shadish et al. (2002) as the most comprehensive resource on the Big 4 validities. We also strongly recommend taking all the coursework that you can in statistics and methods, and recommend firsthand experiences working on research projects to appreciate how each of your decisions in the research process can impact the overall validity of your conclusions. We also encourage you to practice using our Big 4 validity framework in your everyday life. You could begin by creating an exercise like the one in this chapter by picking up your own school newspaper (or, for that matter, any newspaper) and reading it critically. We also encourage you to continue using validity as a framework to judge what you can and cannot conclude from the information around you.

REFERENCES AND FURTHER READINGS

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues & Practice, 17,* 10–22.

Bracht, G. H., & Glass, G. V. (1968). The external validity of experiments. *American Educational Research Journal, 5*, 437–474.

Bryant, F. B. (2000). Assessing the validity of measurement. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 507–619). Washington, DC: American Psychological Association.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Goldstein, I. L. (1991). Training in work organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 507–619). Palo Alto, CA: Consulting Psychologists Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*, 379–387.

Rosnow, R. L., & Rosenthal, R. (1996). *Beginning behavioral research: A conceptual primer* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Sears, D. O. (1986). College sophomores in the laboratory: Influence of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*, 515–530.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, *19*, 405–450.

Wechsler, H., Nelson, T. F., Lee, J. E., Seibring, M., Lewis C., & Keeling, R. P. (2003). Perception and reality: A national evaluation of social norms marketing interventions to reduce college students' heavy alcohol use. *Journal of Studies on Alcohol*, *64*, 484–494.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *45,* 1304–1312.

8

Nonexperimental Research Methods

Stephen F. Davis
Texas Wesleyan University

Randolph Smith's chapter on Experimental Designs (see Chapter 9) indicates that experimental research, in which researchers directly manipulate independent variables (IVs) about which they seek to gather information, record changes in dependent variables (DVs), and control extraneous variables that might produce a flawed experiment, yields the strongest cause-and-effect results. However, not all research projects use the experimental method; in fact, researchers have developed a large number of *nonexperimental* techniques. Although they are classified as nonexperimental, these methods do not necessarily produce inferior results. The choice of a nonexperimental research strategy depends on the type of problem the researcher is interested in, where the research is conducted, and the type of data the researcher gathers. We will examine several types of nonexperimental techniques in this chapter.

OBSERVATIONAL TECHNIQUES

Naturalistic Observation

Researchers who use naturalistic observation seek answers to their research questions by observing behavior in the real world. For example, a researcher who is interested in people's behavior at an ATM machine would locate an ATM machine and observe the behavior of people who use it.

The possibilities for conducting studies using naturalistic observation are limited only by our curiosity and insight into a potential research area. When researchers conduct a naturalistic observation study, they have two goals: (a) to describe behavior as it occurs in the natural setting without the artificiality of the laboratory and (b) to describe the variables that are present and the relations among them. In a naturalistic observation study it is important that the researcher not interfere with or intervene in the behavior being studied. For example, the observer at the ATM machine should be as inconspicuous and unobtrusive as possible; otherwise, the observer might influence or change the behavior of the people using the machine. The presence of an observer is not part of the natural setting for an ATM machine. In this context, researchers often mention the *reactance* or *reactivity effect*, which refers to the biasing of participants' responses because they know they are being observed. For example, if the people using the ATM machine know they are being observed, then they may glance around more often and try to rush their transaction. Clearly, the knowledge that you are participating in an experiment and being observed may result in dramatic changes in your behavior.

Naturalistic observation's main drawback is its inability to yield cause-and-effect results. If it does not lead to the development of cause-and-effect statements, why do researchers use this method? In many instances, the reason is quite obvious: Naturalistic observation may be the only option available if the researcher wants to conduct the project. Psychologists who are interested in reactions to natural disasters such as hurricanes, earthquakes, tornadoes, and fires, ethically cannot create such life-threatening situations just to study behavior; they must gather their data and make their observations under naturally occurring conditions.

Case Studies

When researchers conduct case studies, they observe and record the behavior of one, possibly two, participant(s) over an extended period of time. (Conducting research on the same participant[s] over an extended period of time is typically called a *longitudinal design*.) There are no established guidelines for conducting case studies, hence the procedures used, behaviors recorded, and final reports typically vary considerably. Researchers often use case studies in clinical settings to help formulate hypotheses and ideas for additional research. For example, Paul Broca (1824–1880), a French physician, reported the unusual case of “Tan.” His patient was nicknamed Tan because this word (in addition to obscenities when he was angry or frustrated) was the only word the patient uttered (Howard, 1997). Based on his observations, Broca hypothesized that the brain center that controlled speech production was located in the frontal lobe of the left hemisphere. An autopsy indicated that Broca’s hypothesis was correct; Tan’s left frontal lobe was damaged.

Although case studies can provide interesting and important results, you should keep in mind that their results may be applicable only to the patient whom the researcher observed. Researchers should not generalize beyond the patient they observe. However, if your research goal is to understand the behavior of that one patient, then the case study method certainly can achieve your goal.

Participant Observation

Participant observation is similar to naturalistic observation in that behavior is observed in the real world. Unlike naturalistic observation, however, researchers who use participant observation become part of the group they are studying; the researcher does not try to be unobtrusive. When the research goal is to learn about a specific culture or socioeconomic group, researchers frequently use this technique and typically assume the role of a member of the group in order to make the needed observations.

Glesne (1999) identified two types of participant observation that are related to the degree of participation. The *observer as participant* refers to a researcher who primarily observes a situation but also interacts with the others involved in the study. The *participant as observer* refers to a researcher who becomes a part of the culture by working and interacting extensively with the others. There is a cost-benefit relation with these two approaches. The more immersed you become in a culture, the more you can learn about that culture; however, greater immersion can result in a loss of objectivity.

Participant observation suffers from several additional drawbacks. First, it may take a long time before the participant observer is accepted as a member of the group that is under study. This extended period of time can increase the cost of the project substantially. Additionally, simply being part of the situation does not guarantee that the observer will be accepted. If the observer is not accepted, then the amount of information that the observer can acquire is limited. On the other hand, acceptance of the observer may result in a loss of the observer’s objectivity. Finally, as with other observation methods, participant observation lacks the ability to develop cause-and-effect statements.

Selecting Behaviors and Choosing Recording Techniques

The conduct of an observational study requires a great deal of planning and preparation before the researcher actually attempts to record behavior in the real world. First, researchers must decide which behaviors they will observe. Unless you plan to use video equipment, which may not be possible if you plan to be as unobtrusive as possible, you will not be able to record all behaviors. Second, you will need to decide when and where to record the behaviors in question.

These *where* and *when* questions are deceptively important. Consider the ramifications of observing in one location at one time. Even if the researchers make a large number of observations, their conclusions are limited to just that one situation and one time period. If the researchers are interested in behavior in only one situation occurring at only one specific time, then this limitation is not a problem. However, if the researchers want to generalize their findings to other situations, they will need to make observations in other, different situations. The process of making observations in order to increase the generalizability of the conclusions is called *situation sampling.*

Likewise, even if the researchers make a large number of observations during one time period, they cannot generalize their conclusions to other time periods. Should they wish to generalize their conclusions, the correction for this problem is called *time sampling* and is analogous to situation sampling; the researchers simply make observations during several different time periods. Of course, researchers can, and do, use situation sampling and time sampling simultaneously to determine which of the behaviors under observation are a function of context or time of observation.

Once researchers have decided whether to use time sampling and situation sampling, they must also decide whether they want to describe the behavior(s) of interest qualitatively or quantitatively. If they use the *qualitative approach*, their report likely will consist of a narrative description of the behavior(s) in question. (See Chapter 11 for an in-depth coverage of this methodology.) The researchers should take special care to make sure that the narrative reports use clear, precise language and avoid speculative language. The *quantitative approach* requires a thorough knowledge of the measurement techniques the researcher plans to use and the type(s) of analysis that this measurement system will require.

Interobserver Reliability: The Use of More Than One Observer

Researchers using an observational method must also decide if they will use one or more observers. There are two main reasons for using two or more observers. Observers may disagree about what exactly was observed. Even the use of video recordings does not eliminate the potential need for multiple observers. Someone has to watch the video and evaluate the behaviors.

When researchers use two observers, they can mathematically determine the extent to which their observers agree. This agreement is called *interobserver reliability*. Low interobserver reliability indicates that the observers disagree about the behavior(s) they observed; high interobserver reliability indicates agreement.

What can influence the degree of agreement between the observers? Physical factors such as fatigue boredom, and emotional state, can exert an influence. Likewise, experience observing the behavior in question can influence interobserver reliability. Hence, it is the researchers' task to make sure that both observers are well rested, interested in their task, in good physical and emotional health, and have received the same amount of training and experience with the research project and the behavior(s) under study. The trainer should give clear, precise definitions of the behavior(s) to be observed and should provide concrete examples of positive and negative instances of the behavior(s) in question.

How do researchers measure interobserver reliability? The simplest technique involves determining the number of times the two observers agree and the number of opportunities they had to agree. Once the researcher has determined these numbers, they are used in the following formula:

$$\frac{\text{number of agreements}}{\text{number of opportunities to agree} \times 100 = \text{percent agreement}}$$

This simple formula yields the percent of agreement between the two observers, and many researchers use it as an indication of interobserver reliability. The more accepted measure of interobserver reliability is the square of the correlation coefficient between the two observers multiplied by 100 (i.e., r^2 x 100). The resulting figure gives the percentage of variation that is due to observer agreement; the higher the percentage, the greater the agreement. Although there is not complete agreement, most researchers do agree that good interobserver reliability must involve at least 85 percent agreement.

So far we have seen that observational techniques are among the most widely used nonexperimental techniques available to researchers. In the next section we examine a nonexperimental technique that easily can appear to be an experimental procedure.

EX POST FACTO STUDIES

When researchers study the effects of independent variables (IVs) that they cannot or do not manipulate, they are conducting an *ex post facto study*. In Latin, ex post facto means after the fact; hence, this type of research involves using or examining an IV "after the fact"—it has already varied or changed before the researcher started the project. Thus, the ex post facto study clearly qualifies as a nonexperimental research technique. On the other hand, it does share some similarities with true experiments.

For example, just as in a true experiment, ex post facto research involves recording the behavior(s) of groups of participants and then performing appropriate statistical analyses on these data. For example, Licht (2000) evaluated occupational stress as a function of the sex of the participants and whether they worked for a nonprofit or for-profit organization. She found that (a) nonprofit employees perceived more occupational stress than did for-profit employees, and (b) men had more stress than women. Her research is classified as ex post facto because she had no control over the sex of the participants or the type of organization they worked for. For these reasons, she was unable to make cause-and-effect statements. For example, she did not randomly assign participants to jobs, hence she cannot say that working for the nonprofit organization caused the increased stress levels that she recorded.

Even though ex post facto research is widespread, the use of questionnaires, surveys, tests, and inventories may be even more widespread. We examine these popular techniques in the following sections.

QUESTIONNAIRES AND SURVEYS

Because researchers who use questionnaires or surveys to answer their research questions do not directly manipulate variables, these projects belong in the nonexperimental category. Without question, this type of research is extremely popular. One reason for this popularity is its deceptive simplicity; it seems so easy to conduct research using questionnaires and surveys. All the researcher has to do is secure a group of participants, have them complete the questionnaire or survey, and analyze the results. At that point, the project is completed. As we will see, much more is involved in this type of research than meets the eye.

Descriptive and Analytic Surveys

In order to determine participants' opinions on a topic of interest, researchers may administer a descriptive survey or an analytic survey. Researchers use *descriptive surveys* when they are interested only in determining the participants' characteristics, opinions, or beliefs. The polls that the media conduct on election night to predict the winners/losers are good examples of descriptive surveys; they only seek information (such as frequencies or percentages), not reasons or an understanding of the variables involved. Hence, the goal of the descriptive survey is to describe a specific behavior or characteristic as it exists in a participant sample. These sample results will hopefully be representative of the population from which the sample was drawn.

In contrast, researchers who use an *analytic survey* are interested in determining the reasons for a particular behavior, belief, or opinion—what are the relevant variables and how are they related? For example, the researcher who is interested in *why* voters preferred this or that candidate, in addition to determining the number of people who voted for each candidate, would use an analytic survey to delve into the "why" question.

Creating a Survey or Questionnaire

Because psychology is such a vast and diverse field, researchers may choose to conduct a project and find that a survey or questionnaire dealing with their topic does not exist. In this situation researchers may be tempted to create their own questionnaire. Researchers who find themselves in such a situation should exhaust *all* possible sources for such surveys and questionnaires before attempting to create their own instrument. It is not as simple as it may seem to create a good survey or questionnaire; several time-consuming but crucial steps are involved in developing such an instrument:

- *Preliminary (Pilot) Testing*—The researchers will want to conduct preliminary tests with the new instrument. This preliminary testing is typically called *pilot testing* and may include (a) in-depth interviews with participants to determine the type(s) of questions that would be best for the final version of the instrument, and (b) testing small groups of participants to determine that the instrument is valid (measures what it is supposed to measure) and reliable (yields essentially the same score when administered again).
- *Determine How the Data Will Be Collected*—Researchers need to decide if they will collect their data via mail questionnaires/surveys, face-to-face interviews, telephone interviews, or in a group setting. Who will administer the questionnaires/surveys? What sort of training will the assistants require? Clearly, these are important decisions that can impact the conduct and quality of the research.
- *Determine the Type of Questions*—Researchers can choose among several types of questions: yes-no questions, forced alternative questions, multiple-choice questions, Likert-type questions (selecting a response alternative from a designated scale), and open-ended questions. Without a doubt, the type of questions researchers use directly influence they type of data they gather and the type of analysis they can conduct once they have gathered these data. For example, the yes-no format limits the researcher to counting frequencies and conducting nominal-scale analyses, such as chi square tests.
- *Writing the Questions*—After deciding which type of questions to use, the researcher must actually write the items. Generally speaking, items should be clear, short, and specific; use familiar vocabulary; and be at the reading level of the intended participants. Researchers should be careful to avoid creating questions that would constrain or bias the participants' answers in any way. For example, wording a question in a positive manner likely will result in a large number of positive answers. Also, the researcher should ask other experts in the research area to review the items for bias and unintended meanings. The researcher also will want to conduct pilot tests on the newly created items. During these pilot tests, the researcher will want to discuss the items with the participants; participant insights can be very valuable.
- *Determining Other Information to Obtain*—Researchers also have to decide what other relevant information, typically called *demographic data*, they want to collect. Demographic data may include such items as sex, age, academic major, academic classification, size of home community, and annual income. Lest they forget an important bit of demographic information, researchers should review these items very carefully.
- *Determining Administration Procedures*—Researchers must decide how they will administer the questionnaire or survey and what type of instructions they will provide for the participants. They also need to decide who will administer the informed consent forms and answer participants' questions.

Clearly, if done correctly, the creation of a questionnaire or survey is very time-consuming and involved. Whether or not researchers create their own questionnaires, they must decide how they will administer their instruments. Four of the most popular approaches are mail surveys, personal interviews, telephone interviews, and group administration. We examine these approaches next.

Administering the Questionnaire or Survey

- *Mail Surveys*—Researchers use this method to gather data on just about any topic you can imagine. This approach is popular because researchers can send their questionnaires and surveys to a very large sample and they do not have to be present when the participants complete them. Despite these advantages, mail surveys do have several drawbacks. Researchers never know who completed the surveys and in what order the respondent answered the questions. Also, low return rate is a problem with mail surveys (as few as 25 percent return surveys; researchers consider 50 percent a good return rate). A low return rate and not knowing who completed the surveys can result in a biased sample. Researchers can improve their return rates by (a) including a letter that clearly indicates the nature and importance of the research project, the process used to select the participants, and the researcher's assurance that all responses are confidential, along with a prepaid return envelope in the initial mailing, and (b) sending additional mailings to the respondents.

- *Personal Interviews*—If a trained interviewer administers a survey in a participant's home, response rates go up appreciably; in fact, response rates as high as 90 percent are not uncommon. Compared to mail surveys, face-to-face interviews have several additional benefits: The researcher is sure who answered the questions, the order of answering the questions is controlled, and the interviewer is present to answer any questions. Despite these advantages, personal interviews have several drawbacks: They are costly in terms of both time and money, the potential for interviewer bias exists, and high crime rates are making it more dangerous to conduct such surveys in urban areas.
- *Telephone Interviews*—Telephone interviews allow researchers to overcome many of the problems posed by mail surveys and personal interviews. For example, the interviewer actually talks to the respondent and, therefore, knows who completed the survey and the order in which the respondent answered the questions. Also, the interviewer does not have to enter dangerous, high-crime areas. Additionally, computer technology has made it possible to obtain a large, random sample with ease. In most situations the interviewer can enter the participant's responses as they are made, hence the data are ready for analysis as soon as the interviewer makes the final call. Despite these attractions, telephone interviews have several drawbacks. It is relatively easy for potential respondents to screen calls and, thereby, refuse to participate. Also, telephone interviews preclude the use of visual, nonverbal cues, such as facial expression and body language, by the interviewer. Also, many interviewers and respondents find it difficult to establish rapport on the telephone.
- *Group Administration*—Researchers who use college students as their participants frequently administer their questionnaires or surveys in a group setting. Group administration has several advantages: (a) because several participants are tested at the same time, there are savings of both time and money, (b) the researcher can answer questions, (c) the researcher can evaluate nonverbal cues, and (d) the researcher has some control over the order in which participants complete the questionnaires/surveys and the questions within each testing instrument. The main drawback to group testing appears to be the group setting itself; some participants may be tempted to look on their neighbor's questionnaires/surveys and answer in a similar manner.

INVENTORIES AND TESTS

As we have seen, researchers use questionnaires and surveys to determine and analyze participants' opinions on an issue or topic. On the other hand, researchers use inventories and tests to evaluate a specific characteristic or ability. For example, educational psychologists designed the Scholastic Aptitude Test (SAT) to measure academic potential at the collegiate level. The most popular types of inventories and tests are achievement, aptitude, and personality instruments.

Characteristics of Good Inventories and Tests

Because most researchers use inventories and tests prepared by other professionals, they need to thoroughly examine the instrument before they use it. Some of the questions they will need to answer include the following: How long does it take to administer the inventory or test? Can it be administered to a group? Is the administration procedure complicated? Does a trained professional have to score and interpret the inventory or test? Additionally, researchers will be concerned with the validity and reliability of the inventory or test; these are the two most important characteristics of an inventory or test. When a test is *valid*, it measures what it was designed to measure. A test is *reliable* if repeated administrations yield comparable scores; in other words, the inventory or test is consistent.

Types of Inventories and Tests

- *Achievement Tests*—Researchers use achievement tests when they want to evaluate a person's level of accomplishment or mastery. For example, lawyers must pass the bar examination before they are allowed to practice law. Many achievement tests have a passing score that signifies the minimum level of achievement that a person taking the test must attain.
- *Aptitude Tests*—Researchers use an aptitude test to determine a person's ability or skill for a certain job or in a particular situation. For example, the student who plans to attend graduate school following completion of the baccalaureate degree most likely will be required to take the Graduate Record Examination (GRE). Graduate school admissions committees believe that GRE scores predict an applicant's ability to successfully complete graduate-level courses.
- *Personality Tests or Inventories*—Researchers use a personality inventory or test to measure a specific aspect of a person's motivational state, interpersonal capability, or personality.

SUMMARY

I hope that this chapter has convinced you that psychological researchers do not have to directly manipulate IVs and record subsequent changes in DVs, while extraneous variables are controlled, in order to gather meaningful research data. Nonexperimental methods can produce excellent data; they simply cannot address cause-and-effect relations. Our examination of research methods has only scratched the surface of the available techniques; I encourage you to read further.

REFERENCES AND FURTHER READINGS

Glesne, C. (1999). *Becoming qualitative researchers: An introduction* (2nd ed.). New York: Longman.

Howard, D. (1997). Language in the human brain. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 207–304). Cambridge, MA: MIT Press.

Licht, C. A. (2000). Occupational stress as a function of type of organization and sex of employee. *Psi Chi Journal of Undergraduate Research, 5,* 46–55.

Smith, R. A., & Davis, S. F. (2007). *The psychologist as detective: An introduction to conducting research in psychology* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

9

Experimental Designs

Randolph A. Smith
Lamar University

Knowing about experimental designs is essential if you are planning an experiment. It is highly unlikely that you would decide to set off on a cross-country trip without a road map. Although you might eventually reach your destination, you would likely waste time and money getting there. On the other hand, you might get hopelessly lost and never find your planned stopping point. Likewise, you should never plan an experiment without first considering its design. Although you might be fortunate and plan an experiment that would answer the question that you had in mind, it is much more likely that you would waste your time and end up with a mess on your hands.

PURPOSE OF EXPERIMENTAL DESIGNS

Psychologists use experiments to answer questions about behavior. As a student of psychology, you probably have had many such questions yourself (e.g., why did that person act in that way?). Without research, we are left with no way to answer such questions except by guessing. A well-designed experiment, however, can put us on the road to answering a *causation* question more definitively. The question of causation is an elusive one, but one that scientists seek to answer.

A researcher typically begins with a *hypothesis* about the cause of a particular behavior. A hypothesis is somewhat like an educated guess—based on results from previous research and observations of behavior, a researcher generates a hypothesis about the cause(s) of a behavior or a factor that will affect that behavior. The researcher uses that hypothesized cause or factor as an independent variable (IV) in the experiment. An IV is the variable whose effects you wish to study in your experiment. Of course, to determine the effects of an IV, you must have something with which to measure those effects—in our case, it would be the behavior that the researcher believes the IV will affect. This behavior or outcome variable is known as the dependent variable (DV). So, a researcher plans an experiment to determine the effect of an IV on a DV, which sounds relatively simple. However, complicating this process is another category of variables—extraneous variables. Extraneous variables are variables besides the IV that might have an effect on the behavior in question (the DV). There are probably few, if any, behaviors that are so simple that there is only one causative factor or variable. Thus, when we conduct an experiment, we must take steps to ensure that we are studying only the effect of our IV on the DV and not the effect of any extraneous variable(s). Experimental design is a helpful tool for the researcher in the quest to answer questions of causation in an experiment.

Fortunately, learning about experimental designs is easier than most students think it will be. Once you know a few basics about research and experiments, you can plan an experiment and choose the correct design for it simply by answering a series of questions (see Table 9.1).

Table 9.1 Outline of decisions to choose an experimental design

How many independent variables does your experiment have?
- One
  - How many groups/levels does your independent variable have?
    - Two
      - How did you assign/categorize participants to groups?
        - Randomly (no relationship)
          - Two independent groups design
        - Not randomly (a relationship exists)
          - Two correlated groups design
    - More than two
      - How did you assign/categorize participants to groups?
        - Randomly (no relationship)
          - Multiple independent groups design
        - Not randomly (a relationship exists)
          - Multiple correlated groups design
- Two or more
  - How did you assign/categorize participants to groups?
    - Randomly (no relationship)
      - Factorial independent groups design
    - Not randomly (a relationship exists)
      - Factorial correlated groups design
    - Combination of randomly and not randomly

Factorial mixed groups design

INDEPENDENT VARIABLES

As you can see in Table 9.1, the first question you must answer is "How many independent variables does your experiment have?" For example, suppose people have told you all your life about how important sleep is to good academic performance; thus, you would like to conduct an experiment to investigate the effects of sleep, in particular the notion of "adequate sleep." Sleep would be your IV, so you have only one IV in your experiment. However, your choice of IV should immediately bring to mind some questions about what constitutes adequate sleep. This situation is often the case when you plan an experiment—the IV is somewhat vague, so you must develop an operational definition for the IV. It is important to develop a good, reliable operational definition so that people who read about your experiment will understand what you meant by your IV and could conduct their own experiment using your IV.

Operational Definition

An operational definition is one that entails what you, as the experimenter, have done to cause the IV to occur. As you think about the term "adequate sleep," you can probably think of many different ways to define that term. As the experimenter, you control and are responsible for the way that you define the IV. For example, is adequate sleep six hours of sleep a day? Is it eight hours of sleep a day? Does adequate sleep mean more than simply a certain number of hours of sleep a day? As you can see, deriving an operational definition of your IV of adequate sleep will require some work on your part. Lest you decide that adequate sleep is a bad choice for an IV because it may be difficult to operationally define, you should realize that many potential IVs have similar problems with definitions. Think, for example, about such IVs as hunger, music, therapy, motivation, and so on—all would require a precise operational definition for use in an experiment, just as "adequate sleep" would. Let's assume that you choose to define adequate sleep as eight hours of sleep a day—a fairly standard recommendation that you can find on sleep Web sites and in popular press articles. You now have your IV *and* its operational definition—at least for one group.

Comparison Group

Typically, you, as the experimenter, must also develop an operational definition for a comparison group within your IV. If you wish to study the effects of adequate sleep, you must compare participants who get adequate sleep to some other group (or groups) of participants. Given your interest in adequate sleep, you will probably want to compare your participants to other participants with less-than-adequate sleep (or typical sleep)—but how do you define "less-than-adequate sleep"? As you can probably imagine, you should develop the operational definition of your comparison group relative to the operational definition of your experimental group. Because of much recent evidence about the sleep habits of college students (e.g., http://www.cbsnews.com/stories/2004/04/19/health/main612476.shtml), let's say that you define less-than-adequate sleep as six hours of sleep per night. Just as with the IV, you have a choice of how to define your comparison group.

At this point, students often get confused and decide that they have two IVs: adequate sleep and less-than-adequate sleep. Actually, you simply have one IV (sleep) with two groups or levels (adequate and less-than-adequate) of that IV. As you can tell from examining Table 1, it is critical that you can tell that this example has only one IV. Be sure *not* to confuse IVs with levels of an IV. Levels of an IV are simply the comparison groups of different values (or types) of the same IV.

Measuring the Effect of the IV

When you use an IV in an experiment, you must decide how to measure its effect—in other words, you must have a "yardstick" of some sort to measure whether the IV affected something. For example, in our hypothetical example, you must decide how you would like to assess the effect(s) of adequate and less-than-adequate sleep. Although sleep is the variable of interest to you, you must also have another variable of interest to gauge the effect of sleep. This variable will be your DV. It will be necessary for you to develop an operational definition for your DV, just as you did for the IV. As you can imagine, the DV should be some variable that you can reasonably assume would show the effect of the IV. It would be silly, for instance, to say that you wanted to measure the effect of sleep on eye color—there is no logical link between those two variables. Given that you will probably be conducting an experiment over a short period of time, the DV should be something that sleep might affect over that short duration. Although sleep might affect people's physical health, for example, those effects might take months to achieve. To measure the shorter-term effects of sleep, you decide to give your experimental participants an exam after getting either adequate or less-than-adequate sleep for a full week. After reading that sentence, you may have thought that the experimenter would need to specify an exact exam in order to meet the requirement of an operational definition. If this thought occurred to you, congratulations!

Manipulated Versus Measured IVs

You must also decide how you will present, administer, or measure the IV. As the experimenter, you have the opportunity to be in control of the IV—at least in many cases. If you can actually cause the IV to occur, then you can manipulate the IV. A manipulated IV is your better choice, as you will see. Using your IV in the hypothetical experiment we have been discussing, you could manipulate adequate sleep by controlling how long your participants sleep. You might need to provide them with a location to sleep—perhaps by giving them a bed in a sleep laboratory or some facility where you could monitor how long they sleep. This approach would be more work for you than allowing participants to monitor their own sleep, but it would provide you with more control. On the other hand, sleep also could be a measured IV. Rather than actually *causing* the IV to occur, you could measure the IV in your participants. If you had your participants report their sleeping habits to you, you could measure or classify them as to whether their sleep was adequate or not. Given your operational definition of adequate sleep, however, it might be difficult to find many participants who meet that precise definition. Also, people might not remember exactly how long they slept and simply provide you with their best guess. As you can see, by measuring or classifying your IV, you lose a good bit of control over the situation.

Why would you, as an experimenter, want to measure or classify an IV rather than manipulating it? One potential advantage is that measuring or classifying is often a simpler process. Asking participants about their sleep is far easier than actually manipulating it. But, in the case of adequate sleep, you could use either a manipulated or a measured IV. The main reason for measuring or classifying an IV rather than manipulating it, however, is that some IVs simply cannot be manipulated. Suppose you wanted to deal with personality type as an IV. There is no way to *cause* participants to possess a certain personality; instead, you could only measure personality type and assign participants to groups based on what you measured. Thus, when you choose your IV, you will have to decide whether you can manipulate it or whether you will have to measure or classify it; in some instances, you may have a choice. Let's see why you would want to manipulate your IV, if possible.

Advantage of IV Manipulation

In an ideal world, an experiment is a tightly controlled situation in which the experimenter has total control (or nearly so) over all variables in the environment. Having control allows the experimenter to manipulate the IV, which means that any participant could be a part of any group. Thus, any participant could be in the adequate sleep group *or* the less-than-adequate sleep group because the experimenter controls the participants' sleep. Having control also allows the experimenter to hold other factors in the experiment constant, so that they do not change unless the experimenter wants them to change. Being able to control and manipulate the IV, as well as other variables in the environment, has a major benefit over measuring or classifying the IV: You can attribute causality to the IV. Causality (also known as "cause and effect") allows you to conclude that the IV has actually had an effect.

When you simply classify your IV, you lose the ability to make cause-and-effect statements. If you simply allow experimental participants to report their levels of sleep, you still are able to classify them as getting adequate or less-than-adequate sleep. However, it is possible that an extraneous variable will also become a factor in your research. Imagine, for example, that most of the students who report adequate sleep are freshmen and sophomores, whereas those reporting less-than-adequate sleep are mostly juniors and seniors. In this case, you would have an extraneous variable changing at the same time as your measured IV. Thus, your research would be confounded, which means that you cannot discern the causative variable if your two groups perform differently on the test. It *could* be sleep differences that led to differing scores, but it *could* also be classification differences that influenced the scores—you simply *cannot* tell. On the other hand, if you actually manipulate and control participants' sleep, then you could have classifications spread evenly between your two groups, thus removing classification as a potential extraneous variable. In this manner, you still can draw a causative conclusion about the effects of differing amounts of sleep on exam scores.

NUMBER OF GROUPS/LEVELS

Table 9.1 shows that, with one IV, the second major question to ask is "How many groups or levels does your IV have?" As we have already seen, there *must* be a comparison group that you use to gauge whether your IV has had an effect. If you simply administered your IV to a group of participants and had no group with which to compare them, you would not actually have an experiment. There would simply be no way to determine whether your IV affected behavior in that situation. Thus, the simplest possible experiment has one IV with two groups.

One-IV/Two-Group Experimental Design

This design description specifies the simplest possible experimental design. There is only one IV, with two groups (see Figure 9.1 for a concrete example)—typically, as noted earlier, both a group that experiences the IV and a comparison group; however, there could simply be two comparison groups. To elaborate, there are two typical approaches that experimenters take to using a comparison group. The first approach is to use a control group—a group that does not receive the IV. Thus, if you were interested in the effects of an afterschool tutoring program, the control group would not receive afterschool tutoring, whereas the experimental group would receive the tutoring. The experimental group is simply the group of participants that receives the IV. The second approach to making a comparison group is to give *both* groups the IV, but in differing amounts or types. In our current example, going without sleep for a week is not possible, so you could not use a presence/absence manipulation. Thus, we must use this second approach and vary the *amount* of sleep: 8 hours versus 6 hours. Likewise, if you were conducting a drug study, you could vary the amount of the drug: 5 versus 10 milligrams, for example. On the other hand, if you were studying different therapeutic approaches to depression, you would want to compare different *types* of therapy.

Researchers often use the first approach mentioned in the previous paragraph early in their research with a particular IV—the presence/absence manipulation makes it relatively easy to determine whether the IV has an effect. After researchers know that an IV does have an effect, they can then use the second approach to determine more particular details about the IV's effects (e.g., Does the IV have an effect at low levels? Does one type of IV have a stronger effect than another?). Neither of the two approaches is particularly better than the other one; the specific IV may make the presence/absence manipulation impossible, or the particular research question may dictate one approach over the other. Again, given the particular IV in our example (sleep), we will be using the second approach (varying the amount of IV) from this point on.

Independent Variable (Sleep)

Group 1	Group 2
Adequate sleep (8 hours)	Less-than-adequate sleep (6 hours)

Figure 9.1 Outline of decisions to choose an experimental design.

ASSIGNMENT OF PARTICIPANTS TO GROUPS

The final question that we need to answer (see Table 9.1) revolves around how we form groups of participants. The simplest possible method is to randomly assign the participants to the two groups. Random assignment means that all participants have an equal chance of being assigned to either group. It should be obvious to you that random assignment indicates a manipulated IV rather than a measured IV—if the IV was measured or classified, participants are essentially preassigned to their groups based on the groups (levels) of the IV. When experimenters use random assignment, the resulting groups are called random groups, which means that participants in one group have no relationship to participants in the other group. Because both groups are formed randomly, statistically, they should be equal. For example, if you had 40 students serving as your participants and randomly assigned them to two groups of 20, those two groups should be equated on a variety of variables such as intelligence, personality, motivation, and so on. Beginning with groups that are equal is critical to determining causality attributed to the IV. When an experiment begins with equal groups, and the experimenter treats the two groups identically (control) except for the variation of the IV, then any difference in the groups at the end of the experiment must be due to the manipulated IV. As you can see from Table 9.1, when you have one IV with two groups and assign participants to the groups randomly, you have an experimental design known as a two independent groups design. The "two groups" phrase signifies one IV with two groups, and the term "independent" indicates that the two groups have participants assigned randomly (which makes the groups independent of each other).

Sometimes experimenters worry that random assignment will not result in equal groups. Perhaps the number of participants is quite small, so that random assignment is somewhat less reliable, or maybe the researcher fears unknown variables that might not be equated through random assignment. In such a case, experimenters may choose a nonrandom method of assigning participants to groups so that equality of groups is more certain. There are three such nonrandom methods, all of which result in a relationship between participants in the groups such that they are not independent groups.

1. In *matched pairs* designs, experimenters measure their participants on some relevant variable (e.g., intelligence

for a learning experiment) and then create pairs of participants who are equated on that variable. Thus, as far as the relevant variable is concerned, the participants are the same (i.e., they have the same level of intelligence). For this reason, we speak of a correlation existing between each set of two participants. If we know that age is an important variable in sleep research, then we may want to pair our participants in terms of age—have two 18-year-olds, two 30-year-olds, and so on paired together.

2. In *repeated measures* designs, experimenters measure their participants more than once—under each condition of the IV. For example, in our example using sleep as the IV, each participant would take part in the experiment in both the adequate and less-than-adequate sleep groups (obviously not at the same time)—in essence, each participant serves as his or her own control. Each participant would sleep eight hours a night for some number of days and then experience the DV measurement. The participants would then do the same thing a second time, except for sleeping six hours a night. To prevent the order of treatments from becoming an important factor in the experiment, the experimenter would exert control over the situation and counterbalance the treatments. In this example, counterbalancing means that half of the participants would sleep eight hours a night first, followed by the six hours a night condition. The other half of the participants would get their nights of six- and eight-hour sleep in the opposite order. Thus, the experimenter would have counterbalanced the order of treatments so that any effects of the order would essentially be negated (or cancelled out).

   Repeated measures design is a popular method of creating correlated groups for several reasons: (a) it reduces the need for as many experimental participants because they participate in both conditions of the study, (b) it is a powerful design because participants serve as their own controls, and (c) it is often the case that researchers want to measure participants more than once (e.g., across time) in an experiment. At the same time, however, repeated measures design has the drawback of not being applicable to all research situations. For example, if participants learn something critical to the research during a first measurement (e.g., how to solve a puzzle), then participation in the second phase would not be appropriate (e.g., if it involved solving the same puzzle). In the hypothetical sleep experiment, to use repeated measures, you would have to have a test that you could use more than once *or* two versions of the test that are comparable.

   The fourth edition of the *Publication Manual of the American Psychological Association* (APA; 1994) was the first to recommend that human subjects in an experiment be called "participants" instead, as a more humanizing referent. This practice is followed in this chapter.

3. *Naturally occurring pairs* refers to pairs of research participants who have some natural relationship (often biological in nature). It is much like matching, except that matched pairs are artificially paired, in contrast to these natural pairs. Examples of common natural pairs that researchers might use include twins, siblings, or littermates (for animal studies). For example, in an attempt to determine the effects of heredity and environment on IQ, researchers often work with identical twins separated at birth. Such twins have identical genetics, thus holding heredity constant between them. The twins might, however, differ considerably in their childhood environments—if one twin was reared in a high-quality environment (e.g., attentive parents, good schools, high socioeconomic class) and the other twin grew up in a disadvantaged environment, then we have a naturalistic study that is as close as we can get to an experiment comparing the effects of good and bad environments on IQ. Sometimes the relationship between naturally occurring pairs is not biological in nature; for example, a researcher might use husbands and wives as pairs because of the assumption that marriage would bring about similarities in the couple. The limit to using natural pairs should be fairly obvious: Many times an experimenter wishes to use participants who do not have any natural relationship such as those described; in this case, the experimenter can "create" pairs through the use of matching (described in 1).

Statistical Analysis

Lest you get lost in this forest of experimental design factors so that you cannot see the trees, let me remind you that this discussion has centered only on designs that have one IV and two levels. The reason that all of these factors have fallen under the same main heading is that all such designs require similar statistical analyses. Assuming that your DV yields interval or ratio measurement (see Chapter 6), any one-IV, two-group experimental design will use a *t* test for statistical analysis. The *t* test compares the size of the difference between DV means (in our hypothetical example, exam scores) of the two groups to the underlying variability. If the difference between means is large compared to the variability, then the researcher concludes that the difference is significant (probably not due to chance) and attributes that significance to the IV (see Chapter 6). This conclusion is possible because of the control afforded the experimenter by the experimental design.

The only difference in statistical treatment for the two-group designs depends on the assignment of participants to groups: You would use an independent-groups *t* test to analyze the results from this type of design when you have used random groups. On the other hand, if you used nonrandom groups, the proper statistical analysis would be a correlated-groups *t* test (also known as a *t* test for matched groups or dependent *t* test). This relation between experimental design and statistics is a crucial one—all designs lead logically and invariably to a particular statistical test. This reason is why it is critical for researchers to use a specific design when they develop their experimental ideas.

One-IV/Multiple-Groups Experimental Design

Discussing experimental design is a simpler matter from this point forward—not because the designs themselves are simpler, but because you are now familiar with the important components of experimental designs and their vocabulary. For example, you should be able to read the description of this new design and immediately understand how it differs from the previous design (one IV, two groups). There is still only one IV, but there are now multiple groups based on that IV. If we go back to our hypothetical experiment looking at the effects of sleep, we would have to go beyond having only two groups for adequate and less-than-adequate sleep (see Figure 9.2). As you can see, in this hypothetical example, the researcher has added a group with even less sleep per night to the experiment. This group would sleep far less than is typical for most people and would provide a good contrast for a group that sleeps more than average (eight hours) and a group that sleeps about as much as average college students (six hours). Although Figure 9.2 shows three groups (three levels of the IV), there is no limit to how many levels of the IV you could have in your experiment. Reality, of course, probably does place a limit on how many groups you could have. Realistically, at this point, it may not make sense to add another group to this hypothetical experiment—we have an above-average sleep group, an average sleep group, and a below-average sleep group. Rather than adding more groups to this design, the researcher might conduct this experiment with the three groups and then determine whether further research with more varied sleep groups would be warranted.

Assigning Participants to Groups

All the other concepts that you learned about with regard to the simplest experimental design apply to the multiple-group design also. Chief among those other concepts is the assignment (or classification) of participants to groups. It is still possible to use random groups, typically by assigning each participant to one of the three (or more) groups in some random fashion. In this case, you would end up with three random groups so that participants in Group 1 had no relationship to those in Group 2 or Group 3, and so on.

By the same token, you could also use one of the nonrandom strategies for assigning participants to groups. Because there are now at least three groups, the nonrandom strategies become somewhat more complex. For example, instead of matched pairs, you must use *matched groups*. Instead of simply matching two experimental participants on some variable, you must now match as many participants as you have groups. In our hypothetical example in Figure 9.2, we would need to match three participants in terms of age (or some other relevant variable) and then randomly assign each of them to one of the three sleep groups. If you use *repeated measures* in your experiment, you would need to have participants serve under all conditions in the experiment. Thus, participants would have to participate in more experimental conditions and for a longer period of time. Finally, using natural pairs would have to become *natural groups*. It is likely that this nonrandom strategy becomes least likely as you increase the number of groups. Instead of being able to use twins in your study, you would need triplets, quadruplets, or even quintuplets as your design got larger. Obviously, finding such participants would become exceedingly difficult. To use siblings, you would have to find larger and larger families; also, you could no longer use husbands and wives. The one situation that would still be likely for natural groups is animal littermates, assuming the animal species in question has large litters.

Independent Variable (Sleep)

Group 1	Group 2	Group 3
8 hours	6 hours	4 hours

Figure 9.2 Experimental design with one independent variable and three groups.

Statistical Analysis

Because *t* tests are appropriate for only two groups, a different test is necessary for the multiple-group designs (again, assuming that your DV yielded interval or ratio measurements). In these cases, you would use a one-way analysis of variance (ANOVA) to analyze your data, regardless of the number of groups in your experimental design. The term “one-way” simply refers to the fact that you have only one IV in the design—if you look at Figure 2, you can think of analyzing the data in only one way or one direction: across the design (comparing the three groups to one another). Analysis of variance refers to how the statistical analysis “works.” ANOVA compares the variability among the groups (the levels of the IV) to the variability within the groups (individual differences and error). If the variability due to the IV is large relative to the within-groups variability, then the researcher concludes that the IV has had a significant effect (see Chapter 6). This conclusion tells you that there is a difference somewhere among the means—unlike the *t* test, there are more than two means to compare, so it is not immediately obvious which mean(s) is (are) higher than the other(s). Additional statistical tests would be necessary to determine which groups (levels of the IV) are different from one another.

Two (or More) IV Experimental Designs

Based on Table 9.1, you might think that we have covered the majority of possible experimental designs—only about a quarter of the table is left to cover. However, experimental designs with two or more IVs are so numerous and used so often by experimenters that they are almost the standard type of design used in research. You will also notice from Table 1 that the decision-making process about choosing the appropriate design is simpler for experimental designs with more than one IV—you do not have to ask the question about how many groups or levels each IV has. Any design with multiple IVs falls under the heading of *factorial designs* because IVs are also known as factors; thus, as soon as you plan an experiment with two or more IVs, you know you will be using a factorial design.

Rationale of Factorial Designs

Researchers conduct experiments with multiple IVs because such designs allow a more real-world view of factors that affect behavior (DV). In a single-IV design, you are able to study only one IV's effect on behavior at a time. In the real world, it is not very likely that our behaviors are affected by only one IV. For example, think about your score on the last exam that you took. If you stayed up all night cramming and took the exam with no sleep, you can certainly see how lack of sleep might have affected your score on the exam. However, was that the only factor that had an effect on your exam score? Of course not! What about your level of interest in the class, your physical health status, the number of other exams you had scheduled that day, whether you had a recent fight with your significant other, and so on? All of these factors (and *many* more) could have played a role in how well you performed on the exam. A factorial design allows you to study two (or more) of these factors as IVs simultaneously—it is somewhat similar to conducting two single-IV experiments at the same time.

Benefits of Factorial Experiments

The logic behind a factorial experiment is quite similar to that of a single-IV experiment—with one important change. In a factorial experiment, you want to determine the effects of two (or more) IVs on the same behavior or outcome. You could, of course, accomplish this same objective by conducting two separate single-IV experiments. There are two distinct advantages, however, to conducting a factorial experiment rather than separate single-IV experiments. First, you will typically save time by conducting only one experiment rather than multiple experiments. Second, the major advantage of a factorial experiment is that you get more information than you would from separate single-IV experiments. In both approaches, you would find out how each IV affects the DV—what is known as a main effect. In a factorial design, you will also find out how both IVs *simultaneously* affect the DV—which is known as the interaction of the IVs. An interaction lets you know whether the effect of one of the IVs on the DV depends on the particular levels of the other IV. Again, the possibility of an interaction provides a more real-world view of the behavior in question because most of our behaviors have multiple influences. An example should help clarify the concept of an interaction—a concept that many students find confusing at first.

Beals, Arruda, and Peluso (2002) conducted an experiment dealing with performance on an English proficiency test (the DV). (An aside: Beals and Peluso were undergraduate students at Mercer University when they developed and conducted this experiment; Arruda was a faculty member there.) They hypothesized that the format of the test (multiple choice or short answer—IV1) might affect performance on the test, particularly when taking into account the English proficiency level of the participants (proficient vs. limited proficiency—IV2). Indeed, their results showed a significant interaction (see Figure 9.3) between test format and proficiency level. Graphing an interaction often helps to make sense of the findings. As you can see in Figure 9.3, the two lines are not parallel—this type of picture is often a hint that the interaction of the two IVs is significant. Participants who were proficient in English made similar scores on the test, regardless of whether the test was multiple choice or short answer. Participants with limited English proficiency, however, scored much higher on the multiple-choice test than on the short answer test. From this example, you can tell that the effects of test format *depended* on the participants' proficiency level. Or, you could say that the effects of the participants' proficiency level *depended* on the type of test format. Either way that you describe the data, you can see that you must refer to both IVs *simultaneously* to explain the results of the significant interaction.

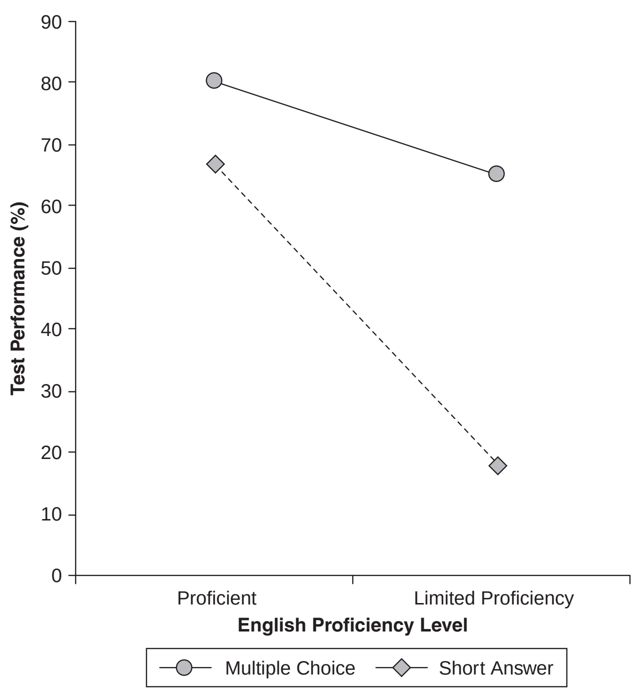

Figure 9.3 Interaction from Beals, Arruda, and Peluso (2002).

Whenever you find a significant interaction, you should interpret it and not the main effects (the individual IVs) because the interaction renders the main effects meaningless. If the effect of one IV depends on the level of the other IV, then it makes no sense to refer to an IV on its own because, even if it showed a significant effect, the effect would not be straightforward. For example, in the Beals et al. (2002) study, it would not make sense to conclude that the multiple-choice test was easier—only the participants who were less proficient in English scored higher on the test.

		Amount of Sleep		
		8 hours	*6 hours*	*4 hours*
Caffeine Ingested	0 milligrams			
	200 milligrams			

Figure 9.4 Experimental design with two independent variables.

Designing Factorial Experiments

Designing factorial experiments is a little like designing two single-IV studies and then combining them into one experiment. The simplest possible factorial design consists of two IVs, each with two levels. The standard way of referring to such a design is a *2 x 2 factorial design*—the number of digits tells you the number of IVs, and the value of each digit tells you how many levels each IV has. Thus, if you had an experiment with two IVs, one with two levels and one with four levels, it would be a *2 x 4 factorial design*.

Returning to the hypothetical sleep experiment, suppose you want to add a second IV because you know it is common for people to drink coffee to “get going” in the morning; this behavior might be even more likely if a person got less-than-adequate sleep. Thus, you might want to determine the joint effect of amount of sleep and amount of caffeine—if less-than-adequate sleep does hamper test performance, might the effect of caffeine counteract that effect? There are almost limitless possibilities for designing this experiment, primarily depending on how many groups and what levels of the IVs you want to use. To keep things relatively simple, suppose that you decide to use only two levels of caffeine for the second IV, which you add to the last version of the sleep experiment (see Figure 9.2). You decide to administer the amount of caffeine (approximately) in two cups of coffee and to compare that to no caffeine. When you combine the two IVs, you end up with the experimental design shown in Figure 4—thus, you have a *2 x 3 factorial design*.

Assigning Participants to Groups

As with the multiple-group design, you have to determine how to assign your participants to groups. You can use random groups or one of the nonrandom strategies discussed earlier. However, because you have multiple IVs, a new possibility arises: You could use random groups for one IV and nonrandom groups for another IV. This possibility is the reason that there are three possible experimental designs listed in Figure 9.1 for a factorial experiment. You could use random groups for both IVs, typically by assigning each participant to one of the groups of combined IVs in some random fashion. In this case, you would end up with random groups throughout the experiment so that participants in any given group had no relationship to those in any other group.

By the same token, you could also use one of the nonrandom strategies for assigning participants to all combined IV groups. Because any factorial design has at least four groups for two IVs, the nonrandom strategies become somewhat more complex. For example, you would have to have participants in all groups matched—or, for repeated measures, participants would have to participate in every possible condition in the entire experiment. Using natural groups becomes highly unlikely in a factorial design.

The third possibility, mixed groups, entails one IV with random groups and one IV with nonrandom groups. This option is a popular one, as researchers often like to measure random groups (IV1) over time (repeated measures—IV2).

Statistical Analysis

Just as with the multiple-group designs, you will use ANOVA to analyze the results from factorial designs (assuming, of course, that your DV yielded interval or ratio measurements). However, for such designs you would use a factorial ANOVA to analyze your data, regardless of the number of IVs in your experimental design. You might hear the term “two-way ANOVA,” which refers to the fact that you have two IVs (often referred to as A and B) in the design (or “three-way” for three IVs: A, B, and C). In such a case, ANOVA still compares variability due to IVs to the variability within groups (still individual differences and error). In a two-way factorial ANOVA, however, you get results for each IV plus the interaction of the two IVs (A, B, and AB). As you add more IVs, not only do you get results for more main effects, but the number of interactions also increases rapidly. For three IVs, you would get results for A, B, C, AB (two-way interaction between Factors A and B), AC, BC, and ABC (three-way interaction among all IVs). With four IVs, you would find results for A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, and ABCD (four-way interaction). As you might guess, it is rare that researchers use designs larger than three or four IVs.

VARIATIONS ON EXPERIMENTAL DESIGNS

Researchers often want to compare participants of different ages to get a glimpse at how behavior may vary or change over the lifespan. There are two approaches to measuring these differences—each is associated with a particular type of experimental design.

A powerful method for looking at changes during the lifespan is longitudinal research. In this approach, a researcher studies a group of participants over an extended period of time, measuring them repeatedly throughout that time. One of the best-known longitudinal studies is that of Lewis Terman, who began studying more than 1,500 gifted children (IQs above 140) in the 1920s. Although Terman died in the late 1950s, the study has continued and will not end until all of the original participants have died (http://www.hoagiesgifted.org/eric/faq/gt-long.html). Because the researcher measures the same participants repeatedly, a longitudinal study employs nonrandom groups (repeated measures). Thus, a researcher conducting a longitudinal study would use one of the correlated groups designs in Figure 9.1.

As you can imagine, there are many barriers to conducting a longitudinal study. Two major problems are the long time involved in such a study and the difficulty in tracking participants over many years. To alleviate these problems, researchers developed cross-sectional studies. In a cross-sectional study, a researcher measures participants from different age groups at about the same, limited time. For example, you may have heard of the Nielsen TV ratings—they measure TV viewing habits of different age groups at the same time to determine how viewing habits vary by age. Although the basic purpose is the same as in longitudinal research, cross-sectional studies enable researchers to gather the data in a much shorter time period and to avoid the problem of tracking participants over time. Because the researcher measures different, unrelated groups of participants, cross-sectional research typically uses random groups and one of the random groups designs from Figure 9.1. If, however, a researcher wanted to ensure the equality of the groups before measuring them, matching might be possible. For example, if you are collecting TV ratings for different age groups but you worry that socioeconomic status might play a role in TV viewing habits, then you might want to use different age groups who are matched on socioeconomic status. In this case, you would use a correlated groups design from Figure 9.1.

SUMMARY

Knowing the basics of experimental design is crucial to planning an experiment. If you simply “threw an experiment together” that did not fit a known experimental design, then you would have no way of analyzing the data and drawing conclusions from your research. Experimental designs give researchers a way of asking and answering questions about their area of interest—human and animal behavior, for psychologists.

REFERENCES AND FURTHER READINGS

Beals, J., Arruda, J. E., & Peluso, J. P. (2002). The effects of language proficiency on test performance. *Psi Chi Journal of Undergraduate Research, 7,* 155–161.

10

Single-Subject Designs

Bryan K. Saville
James Madison University

Since the emergence of psychology as an independent discipline in the late 1800s, research methods have played an ever-important role in the evolution of psychology. Early on, the adoption of research methods from the more established field of physiology helped distinguish psychology as its own separate discipline, distinct from philosophy and ultimately interested in gaining a reputation as a field that took its science seriously (Goodwin, 2005). In fact, if one were to travel back in time to the days of the earliest psychologists, one would notice that the field of psychology *was* a purely research-oriented discipline, a discipline whose primary goal was to elucidate the workings of the human mind.

Much has changed since those early days of psychology. Although many psychological researchers continue to examine topics that were of interest to early psychologists (e.g., memory, perception, consciousness), the field has expanded considerably. For example, whereas topics such as cognitive neuroscience and health psychology were well off the psychological map in the early to mid 1900s, they now demand the time and effort of many psychologists. Similarly, although clinical psychology was but a faint light on the distant horizon in the early 1900s, it now constitutes the single largest of subdiscipline in the field of psychology.

In fact, psychology has exploded to such an extent in the last 50 years that some psychologists would argue that now it consists of nothing more than a large number of disjointed subareas, many of which have little relation to one another (e.g., Kimble, 1999; Staats, 1999; Sternberg, 2005). Whether this point is true is arguably debatable—a debate I do not wish to enter at this time. Nevertheless, the rapid expansion of psychology has resulted in the rigorous study of a wide variety of topics that researchers approach from many different theoretical perspectives. Yet one common thread seems to bind these researchers together. Regardless of their vastly different educational backgrounds or widely divergent theoretical perspectives, psychologists tend to agree that research methods hold a venerated spot in psychology. As Stanovich (2001) suggested, research methods serve as the glue that holds together a discipline consisting of many different subdisciplines.

The most common approach to conducting psychological research—regardless of the specific topic one is studying or the specific theoretical orientation one takes when interpreting observational data—entails randomly assigning large numbers of subjects to different groups, manipulating an independent variable, controlling possible confounding variables, and comparing those groups to see if they differ on some dependent measure of interest. If the groups differ significantly at the end of the study, as shown by an appropriate statistical test, the researcher concludes that the independent variable likely *caused* the observed changes in the dependent variable. This approach, known as the *large-N* or *groups approach* (see Campbell & Stanley, 1963; Cook & Campbell, 1979), has been the dominant approach to conducting psychological research for nearly a century. A perusal of most psychology journals will show that researchers have preferred, and continue to prefer, this approach to learning about the phenomena they study (e.g., Baron & Perone, 1998; Blampied, 1999; Boring, 1954; Saville & Buskist, 2003). In addition, a large majority of research methods textbooks contain extensive discussions of how to conduct research using the groups-oriented approach (e.g., Kerlinger & Lee, 2000; Mitchell & Jolley, 2004). Thus, established researchers, whether implicitly or explicitly, continue to promote the idea that large-N group designs constitute the sine qua non for conducting psychological research. Moreover, students new to the field of psychology frequently come to believe that large-N designs are the only way to determine if there are meaningful relations among the variables they study.

Although large-N designs constitute the most popular means of studying psychological phenomena, another approach to conducting psychological research has a storied place in the history of psychology. *Single-subject*, or *small-N*,[1] *designs* comprise yet another category of research designs that allows researchers to examine functional relations between variables. These designs, most commonly associated with behavior analysis, have become increasingly popular in psychology, most notably in areas of applied psychology where practitioners typically work with small numbers of people (Robinson & Foster, 1979; Saville & Buskist, 2003). The purpose of this chapter is to discuss in more detail the use of single-subject designs in psychological research. Specifically, I will discuss (a) the history of single-subject designs in psychology, (b) problems with large-N designs and why single-subject designs have regained some popularity in the past 30 to 40 years, (c) characteristics shared by different single-subject designs, and (d) some of the most common single-subject designs.

HISTORY OF SINGLE-SUBJECT DESIGNS

Although the use of large-N designs has been a veritable mainstay of psychological research since the early 20th century, research in psychology has not always followed this well-trodden path. In fact, some of the earliest psychological researchers conducted studies in which a small number of subjects responded under different treatment conditions, and the researchers measured the extent to which their subjects' responses varied as a function of these manipulations. Gustav Fechner, for example, considered by many to be one of the founders of modern-day experimental psychology (Boring, 1957; Goodwin, 2005), conducted his early psychophysical research with just a handful of subjects. Similarly, Wilhelm Wundt, often called the "Father of Modern Psychology" because of his willingness to assert that psychology was its own discipline, typically asked individual subjects to introspect on their immediate conscious experience.

Around the same time Wundt and his students were attempting to examine the structure of mind, Hermann Ebbinghaus was conducting his seminal studies on memory. Unlike Fechner and Wundt, however, Ebbinghaus served as his own research subject. Consequently, the well-known retention curves that one can find in most introductory and cognitive psychology textbooks came from Ebbinghaus' measurement of his own memory. Likewise, early learning theorists, including Edward L. Thorndike and Ivan P. Pavlov, used a small number of cats and dogs, respectively, to examine what many would come to know as instrumental (i.e., operant) conditioning and Pavlovian (i.e., classical) conditioning. In essence, then, the use of single-subject designs characterized the early days of experimental psychology.

Beginning in the early 1900s, however, the use of single-subject designs began to wane, and a new approach to conducting psychological research emerged. Influenced by the early statistical work of Sir Ronald A. Fisher (1925) and others, researchers soon found themselves developing studies that utilized large numbers of subjects and statistical analysis to identify significant between-groups differences in their dependent variables. By the mid-20th century, psychology was no longer a discipline that emphasized the use of small numbers of subjects under tightly controlled experimental settings. Instead, psychology was evolving into a discipline that focused on average group differences.

Although the use of many subjects and statistical analyses became a staple of psychological research in the mid-20th century—and has continued relatively unabated to this day—a handful of researchers persisted in their use of the single-subject designs that had been the hallmark of psychological research in the days of Wundt, Ebbinghaus, and Pavlov. The most ardent supporter of single-subject methodology was a relative newcomer to the field of psychology, a freshly minted assistant professor at the University of Minnesota who would soon become one of the most eminent psychologists in history. Burrhus Frederic (B. F.) Skinner, who in his 1938 treatise *The Behavior of Organisms* first outlined his "radical behavioristic" approach to the scientific study of behavior, suggested that the intensive study of a single organism would yield more information about the causes of behavior than a briefer comparison of large numbers of subjects. Skinner's operant psychology eventually gave birth to a small but enthusiastic faction of experimental psychologists, among them Murray Sidman, whose influential *Tactics of Scientific Research* (1960) further outlined the single-subject approach to conducting research in psychology. Moreover, although large-N designs had supplanted the single-subject approach years earlier, applied behavior

[1] Although "single-subject" and "small-N" are essentially interchangeable, Saville and Buskist (2003) suggested that "single-subject" is a misnomer because researchers rarely analyze the behavior of a single subject. Nevertheless, because "single-subject" is more common in the research literature than "small-N," I will use the term "single-subject design" throughout the remainder of this chapter.

analysts continued to follow in Skinner's early footsteps in their attempts to modify socially important behaviors (Baer, Wolf, & Risley, 1968). As Skinner (1966) later stated, "Operant methods make their own use of Grand Numbers: instead of studying a thousand rats for one hour each, or a hundred rats for ten hours each, the investigator is likely to study one rat for a thousand hours" (p. 21). Although Skinner enthusiastically exhorted the benefits of single-subject designs, his arguments did little to diminish the rise of large-N designs in psychology.

THE COMEBACK OF SINGLE-SUBJECT DESIGNS

Although large-N group designs surfaced in the 1920s and eventually became the dominant methodological approach to conducting research in psychology, the field, in fact, saw a modest reemergence of single-subject designs in the 1950s and 1960s. There were at least two primary reasons for the reemergence of single-subject designs during this time (see Robinson & Foster, 1979; Saville & Buskist, 2003).

First, with the rise of applied psychology in the mid-20th century came a need for research designs that would allow practitioners to accommodate smaller numbers of subjects but still see the efficacy of their interventions. Whereas many psychological researchers—especially those in academic or government settings who had easy access to a pool of research subjects—had the luxury of using large numbers of subjects in their studies, applied practitioners typically had neither the time nor the resources to conduct traditional large-N studies. Moreover, for reasons that seem fairly obvious, applied practitioners, especially those whose primary goal was to "hang out a shingle" and provide clinical services to needy clients, were probably reluctant to withhold treatment (due to the lack of a control group) from individuals who had sought out their services. Thus, single-subject designs provided applied psychologists with a way of determining the efficacy of their treatments without using large numbers of subjects or withholding treatment from vulnerable individuals.

Second, along with the ascension of Skinner's operant psychology came an increase in the number of psychologists who bought into his philosophy regarding the use of single-subject designs. As the topics in which these researchers were interested grew more complex, there came a need for more sophisticated research designs that could accommodate the increasingly multifaceted questions these researchers were attempting to answer. As a result, many behavior analysts in the 1960s and 1970s spent a good amount of time developing new and improved single-subject designs. In addition, as single-subject designs became more sophisticated, growing numbers of researchers outside of behavior analysis began to identify ways in which they could use these designs to learn more about the varied phenomena they studied. As a result, one can now find researchers in many different quarters of psychology using various single-subject designs to examine their subject matter.

PROBLEMS WITH LARGE-N DESIGNS

Thus, single-subject designs became more prevalent in psychology during the 1950s and 1960s because of the need for practical research designs that applied psychologists could use with small numbers of subjects and because researchers were developing more sophisticated single-subject designs. In addition, some researchers have also become increasingly reliant on single-subject designs because of flaws inherent in the more popular large-N group designs. Barlow and Hersen (1984) nicely outlined some of the problems with large-N designs (see also Johnston & Pennypacker, 1993; Sidman, 1960).

First, the large-N approach becomes potentially problematic when research questions revolve around the efficacy of a given clinical intervention. For example, imagine a large-N researcher who wished to examine the extent to which a new treatment positively impacted the severity of depression in young adults. Most often, large-N researchers would obtain a sample of clinically depressed subjects and then examine the efficacy of a certain treatment by randomly assigning subjects to the treatment condition or a control (i.e., no treatment) condition. In such cases, a question arises regarding the ethicality of withholding treatment from a group of subjects—in this case, a group of clinically depressed individuals—who may be in dire need of treatment. By withholding treatment in such a study, subjects in the control condition may, at best, not get any better, or may, at worst, continue to suffer. Although some clinical researchers attempt to rectify this issue by promising treatment to subjects in the control condition once the study is complete (Miller, 2003), this strategy does not negate the fact that some subjects in the control condition may suffer during the course of the study.

Second, the use of large-N designs may raise practical issues with regard to obtaining research subjects. Most often, the use of large-N designs requires, for statistical reasons, that researchers strive to include a minimum of 20 to 30 subjects in each group in their study. Thus, because most large-N designs include, at the very least, an experimental group and a control group, the number of subjects in such a study tends to be relatively large. Unfortunately, depending on the type of study, obtaining a sample of this size may be difficult. Assume, for instance, that a researcher is attempting to determine whether a new treatment reduces the severity of symptoms associated with schizophrenia. Because only a small percentage of individuals are schizophrenic, it may be difficult to obtain a large enough number of subjects so that the outcomes are statistically valid. Even if a researcher is able to obtain a large enough number of subjects to create a suitably valid study, other practical issues, such as paying subjects or finding time to collect data from a large sample of individuals, may create practical limitations that interfere with the researcher's ability to conduct a scientifically sound study.

A third problem arising from the use of large-N designs concerns the averaging of group data. The most common practice in group research entails collecting data from a large number of subjects and then providing descriptive statistics for each group that typically include a measure of central tendency (e.g., mean) and a measure of variability (e.g., standard deviation). Based on how much the group distributions overlap, an inferential statistic (e.g., a *t* test or ANOVA) provides information regarding the likelihood that the groups are significantly different from one another.

Unfortunately, a measure of central tendency, although accurately representing the "average" response of the group, may not represent any single individual within that group (Sidman, 1960). For example, imagine a study on exercise and weight loss in which a researcher randomly assigned 100 subjects to either an exercise or a control condition. At the end of the study, the researcher finds that subjects in the exercise condition lost an average of 25 pounds and subjects in the control condition lost an average of 15 pounds, an outcome that might produce a statistically significant result. There is at least one potentially sizeable problem with these hypothetical results. Although subjects in the exercise condition lost more weight *on average* than subjects in the control condition, it is possible that no single subject lost either of these amounts of weight. Moreover, some subjects in the control condition may have actually lost *more* than 25 pounds, and some subjects in the exercise condition may have actually lost *fewer* than 15 pounds. Thus, although one might conclude that "exercise had a positive effect on weight loss," closer analysis of the data shows that group averaging may have obscured some important observations about individual subjects—specifically, that other uncontrolled factors likely also influenced weight loss in at least some, if not many, of the subjects.

Averaging data also may conceal trends in a data set that might otherwise provide important information regarding the effects of an independent variable. Consider, once again, a study in which a researcher measured over eight weeks how diet or exercise affected weight loss in a sample of randomly selected and assigned college students. At the end of the study, the researcher finds that subjects in both conditions lost approximately 10 pounds, on average. Based on this observation, the researcher might conclude that diet and exercise are not significantly different with regard to how much weight loss they produce. Yet closer analysis of the data might reveal additional information that could be important to consider when making statements about the effects of diet or exercise on weight loss. For example, imagine that subjects in the diet condition lost the majority of their weight in the first two weeks of the study but lost little weight after that. In contrast, subjects in the exercise condition lost weight more slowly to begin with, but their weight loss was relatively consistent across all eight weeks of the study. For individuals who are hoping to lose weight and keep it off, this information might be especially important. Without examining individual data more closely, however, the researcher might overlook this important information.

Fourth, the way researchers typically conduct group research often creates problems when it comes to making statements regarding the generality of their findings. Most often, statements regarding the generality of a set of outcomes are based on the assumption that a sample of randomly selected subjects adequately represents some larger population. Thus, findings observed in a certain study are supposed to be an adequate indicator of what one would observe if the study were conducted with different individuals from the same population. Take, for example, a study on the factors that affect vocal tics in adults. One might assume that a randomly selected sample of subjects would represent all adults with vocal tics in every possible way (e.g., age, race, gender). In reality, however, researchers often do not have access to a truly random sample of subjects (e.g., Baron & Perone, 1998; Johnston & Pennypacker, 1993). Consequently, they end up recruiting subjects who are often unique in some way (e.g., introductory psychology students; see, e.g., Sears, 1986), which may severely limit the generality of their findings. In addition, even if researchers do have access to a truly random group of subjects, they often attempt to control individual differences as much as possible in an effort to increase internal validity and, consequently, the likelihood that they will observe a statistically significant outcome. For example, rather than using a truly random sample of subjects with a wide range of individual differences, researchers may limit their sample to subjects who are of a specific age, gender, race, background, and so on. Consequently, any significant findings may only hold for a very limited number of individuals with vocal tics and not for many in the greater population.

As Baron and Perone (1998) stated, "The emphasis on samples and populations is borrowed from the logic of inferential statistical analysis, the goal of which is to infer *population parameters* [italics added] (usually means and variances) from the scores of a representative sample of subjects" (p. 64). Unfortunately, many researchers who use large-N designs and inferential statistics make two important mistakes while interpreting the results of their studies. First, they make generalizations about their data that go beyond the populations from which they drew their relatively narrow samples; second, they make generalizations from a sample of subjects to *individuals* in a greater population. Ultimately, both of these practices greatly misrepresent the conclusions that one can make based on the outcomes of a single study.

CHARACTERISTICS OF SINGLE-SUBJECT DESIGNS

A perusal of any book devoted to single-subject methodology will reveal that there are, in fact, many different types of single-subject designs (e.g., Barlow & Hersen, 1984; Cooper, Heron, & Heward, 1987; Johnston & Pennypacker, 1993). Nevertheless, all of these designs, as strikingly different as they may seem, share certain characteristics. These characteristics include (a) the use of a small number of subjects; (b) exposure to all treatment conditions (i.e., all levels of the independent variable); (c) repeated measurement of the dependent variable under each treatment condition; (d) visual, rather than statistical, analysis of data; and (e) the use of direct and systematic replication in an attempt to make statements regarding the generality of findings (Saville & Buskist, 2003).

NUMBER OF SUBJECTS

Whereas traditional large-N group designs typically include a large number of subjects, single-subject designs utilize a much smaller number of subjects. Although the number of subjects in these studies may vary widely, most researchers who use single-subject designs employ no more than 10 to 12 subjects (and sometimes only 1 subject). For example, recent issues of the *Journal of the Experimental Analysis of Behavior*, the flagship journal for basic single-subject research, contain studies that most often used no more than 10 subjects, although a small number of studies used 20 or more subjects. In contrast, a recent issue of the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, a leading outlet for basic experimental research conducted in the large-N tradition, contains studies that used no fewer than 16 subjects and sometimes as many as several hundred subjects. Whereas the use of large numbers of subjects in traditional group research rests on certain statistical assumptions, single-subject researchers believe that functional relations can be gleaned using only 1 subject responding under different treatment conditions.

EXPOSURE TO ALL TREATMENT CONDITIONS

In traditional large-N group designs, subjects typically receive exposure to only one level of the independent variable. For example, in one of her early studies on the malleability of memory, Elizabeth Loftus (1975) assigned subjects to different treatment conditions in which they did or did not receive "leading" questions about a video they had observed (e.g., "Did you see *the* school bus in the film?" or "Did you see *a* school bus?"). Loftus then examined the extent to which subjects in each group later remembered seeing a bus in the video. She observed that subjects given the leading questions were more likely to report seeing a bus than subjects who did not receive the leading questions.

In the large majority of studies conducted in psychology (e.g., Baron & Perone, 1998; Blampied, 1999; Saville & Buskist, 2003), a researcher randomly assigns subjects to groups, or treatment conditions, and then examines the extent to which the "average" subject in the conditions differs after exposure to the different treatments. Because random assignment means that the outcome of a coin toss, for example, determines whether subjects will be in one group *or* the other (e.g., "heads," you're in Group 1; "tails," you're in Group 2), each subject ends up experiencing only one of the treatment conditions.

In contrast, subjects in studies that utilize single-subject designs typically experience all levels of the independent variable, or all treatment conditions. Take, for example, an early study by Hall, Lund, and Jackson (1968) in which they examined the effects of social reinforcement on studying behavior in elementary school students. Hall et al. first measured "on-task" studying behavior in six subjects under a baseline, or control, condition (i.e., in the absence of social reinforcement for appropriate studying). In the subsequent treatment condition, Hall et al. taught the teachers to deliver social reinforcement (e.g., a pat on the shoulder) to the students when they were on task and measured the accompanying changes in studying. Hall et al. then reintroduced the baseline condition and examined if on-task studying returned to its initial baseline levels. Finally, Hall et al. introduced the social reinforcement condition one last time, examined its effects, and observed that the contingent presentation of social reinforcement increased studying behavior relative to baseline levels.

Whereas a more traditional large-N groups approach to examining the effects of social reinforcement on studying might have entailed the random assignment of a large number of students to either a "social reinforcement" condition or a control condition, the subjects in Hall et al.'s study experienced each treatment condition twice: baseline, social reinforcement, baseline, and social reinforcement. In this way, each subject served as his or her own "control," against which Hall and his colleagues compared the effects of social reinforcement. Thus, regardless of whether each subject was on task a lot or very little during baseline, Hall et al. were able to examine the extent to which social reinforcement affected the unique studying behavior of each individual child.

MEASUREMENT OF THE DEPENDENT VARIABLE

In most large-N designs, researchers typically measure the dependent variable of interest no more than once under each treatment condition. To illustrate, consider a series of studies on stereotype threat by Steele and Aronson (1995). In their studies, Steele and Aronson asked white and black college students to take a 30-item test consisting of questions taken from a Graduate Record Examination study guide. Students did so under one of three treatment conditions: (a) a "diagnostic" condition, in which Steele and Aronson presented the questions as a valid test of intellectual ability; or (b) one of two slightly different "nondiagnostic" conditions, in which participants believed they were simply solving verbal problems and that the test did not involve a real test of intellectual ability.

Steele and Aronson (1995) found that black students in the diagnostic condition performed more poorly than black students in either of the nondiagnostic conditions. They also observed that black students in the diagnostic condition performed more poorly on the quiz than white students in the same condition but not differently from white students when they believed the test was not a true test of intellectual ability. Notice that Steele and Aronson measured their dependent variable only once for each subject. Once they had obtained a test score from each student, they calculated an average score for subjects in each treatment condition and then compared the outcomes statistically.

In contrast, single-subject designs entail the repeated measurement of the dependent variable under each treatment condition. More specifically, single-subject researchers usually measure their dependent variables under each baseline and treatment condition until the behavior of interest has stabilized. Although there are no formal rules for determining whether a particular response is stable, most single-subject researchers agree that at least three data points are needed to determine if a response is stable or if a particular behavioral trend exists (Johnston & Pennypacker, 1963; Sidman, 1960).

For example, consider a study by Smith and Ward (2006) in which they examined the effects of public posting (i.e., posting public information on a player's performance) and goal setting on several different skills in three college football players. Smith and Ward first examined the percentage of times that each player performed the skills (e.g., blocking, running a correct route) correctly under a baseline condition that consisted of 16 practice sessions and two games. Then, Smith and Ward introduced public posting and examined how it affected the players' performances over the course of four more practice sessions and one game. Following another four-session (three practices and one game) baseline condition, Smith and Ward measured their subjects' performances under a "goal setting" condition (six practices and two games), another baseline condition (three practices and one game), and a "public posting and goal setting" condition (six practices and two games). In total, Smith and Ward measured each skill a total of 47 times under the different conditions and found that public posting and goal setting increased the extent to which their subjects performed the different skills correctly.

There are at least three reasons why single-subject researchers prefer to measure their dependent variables multiple times under each condition (Saville & Buskist, 2003). First, repeated measurement may allow a researcher to identify and subsequently control further extraneous factors that may be introducing unwanted variability into the study. If, for example, a subject's responding is quite variable during a particular phase of a study, a researcher may try to identify and remove any additional factors that may be contributing to the observed variability. By doing so, a researcher can increase the likelihood that any changes in behavior are, in fact, a function of independent variable manipulations and not some other uncontrolled factor. In contrast, researchers who use large-N group designs often "weed out" such unwanted variability by using (a) random assignment, which, they assume, will equally distribute unwanted factors to each treatment condition; and (b) statistical analysis, which will tell them how much of the variability in their study is due to random, extraneous factors and how much is due to the independent variable.

Second, single-subject researchers prefer repeated measurements because it affords them the opportunity to interact very closely with their data and manipulate an independent variable at a moment's notice if they wish to do so. For example, imagine a team of researchers studying the effects of a new treatment of self-injurious behavior (SIB) in mentally retarded adults. If our research team introduced a treatment and subsequently realized that the treatment was, in fact, exacerbating the frequency of SIB, they could modify the treatment on a whim in hopes of reducing the potentially damaging behavior. Thus, repeated measurement of the dependent variable allows single-subject researchers to examine the effects of one or more independent variables *while* they are in the process of collecting their data. In contrast, large-N researchers typically measure their dependent variables only once. Consequently, they can examine whether their groups are significantly different (i.e., whether their independent variables had an effect) only *after* they have completed their data collection. Perone (1999) suggested that the traditional practice of measuring a dependent variable only once keeps researchers from having the intensive interaction with their data that they need to obtain a true understanding of the relation between their independent and dependent variables.

Finally, the introduction of a new treatment condition may produce reactivity on the part of the subjects. Consequently, a subject's first response may not be the response that a researcher would observe once the novelty of the new treatment wore off. For example, consider a study in which a researcher wanted to know whether the contingent presentation of movie gift certificates increased the number of minutes that college students exercised each day. The introduction of such a novel treatment might initially result in an increase in the amount of time that subjects exercise during the first week or two. However, with repeated measurement, our researcher might find that the gift certificates do not have a long-lasting effect and that exercise might ultimately fall to a level that is considerably lower than what she observed during the first two weeks. Only with repeated measurement would our researcher be able to determine the long-term effects of her intervention. If, however, our researcher had measured the dependent variable only once or twice, he or she may have come to the mistaken conclusion that the contingent presentation of movie gift certificates may be an effective way to increase exercise in college students.

ANALYSIS OF DATA

Psychological researchers have a long and storied tradition of using statistical procedures to analyze the outcomes of their studies. Although early research in the history of psychology did not employ common statistical analyses (simply because such analyses had yet to be developed), the introduction of statistical tools such as the Pearson correlation coefficient, the *t* test, and analysis of variance (ANOVA) gave psychological researchers new and exciting ways to analyze their data. Specifically, inferential statistics provided researchers with a quantitative means of determining whether their groups were meaningfully different.

Thus, after researchers have collected their data using traditional large-N methods, the use of hypothesis testing and inferential statistics allows them to determine whether any difference in group means was due to random, uncontrolled factors or whether the difference was due to the effects of the independent variable. In addition, the relatively recent introduction of new statistical concepts (e.g., power, effect size) has allowed researchers to be even more certain of the statements they make regarding relations between their variables.

In contrast, single-subject researchers typically analyze their data using visual analysis. More specifically, they often plot their data on line graphs and analyze whether there seem to be meaningful differences in responding under different treatment conditions. For example, imagine once again that a researcher was examining whether a new treatment reduced SIB in a developmentally disabled child. After measuring the child's behavior in the absence of the treatment (i.e., baseline), the researcher might introduce the treatment and measure the instances of SIB every day for the next week. After the number of SIB episodes had stabilized, the researcher might then remove the treatment (i.e., an A-B-A design; see below) and examine whether there seem to be noticeable differences in steady-state levels of responding under the baseline and treatment conditions (see Figure 10.1a). If there is little overlap in the data points under each condition, our researcher can conclude with some confidence (assuming other extraneous factors have been controlled) that the treatment was likely responsible for the observed differences. Depending on the study and what type of behavior is being measured, researchers might also look for different trends in responding (see Figure 10.1b). Finally, although there are no explicitly stated rules for determining whether a response has stabilized or whether there are significant differences across treatment conditions (Sidman, 1960), most single-subject researchers tend to agree that a visually significant difference likely denotes a statistically significant difference and that visual analysis provides a valid method for analyzing functional relations between variables.

Even though visual analysis has been the method of choice for analyzing data in single-subject designs, it is becoming increasingly common for single-subject researchers to use statistical tools. Consider, for example, recent research on delay discounting, which examines the extent to which subjects discount, or devalue, rewards as a function of the delay in their receipt. Most often,

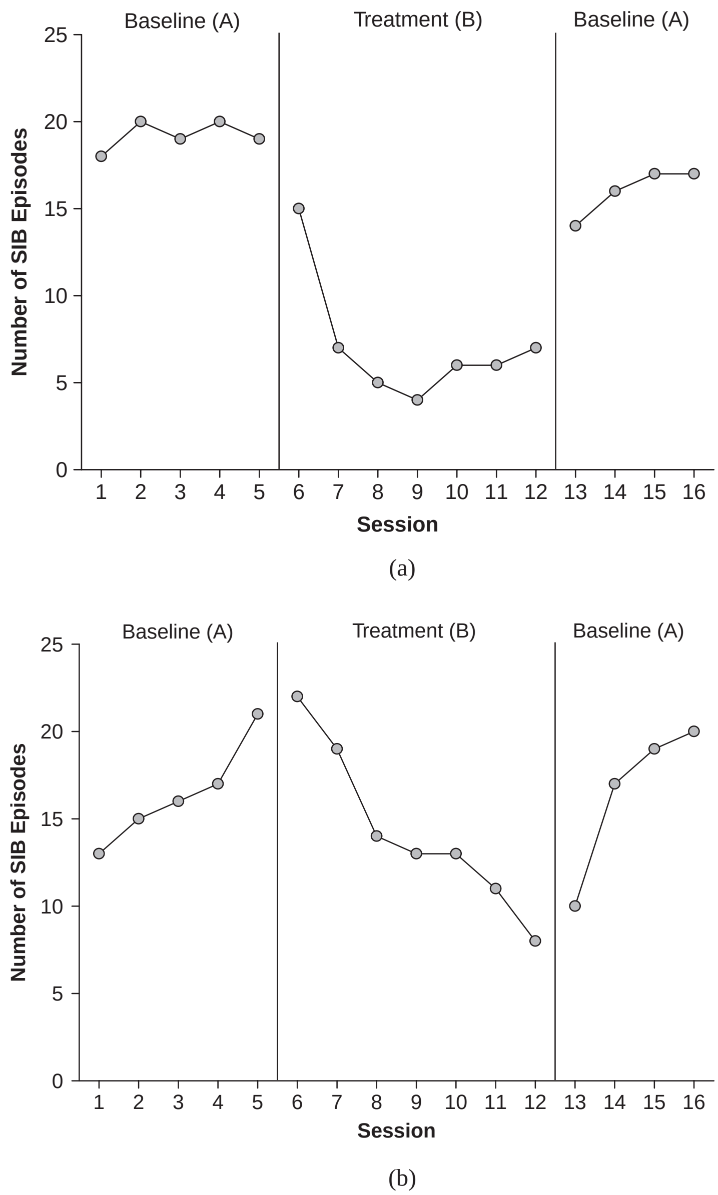

Figure 10.1 The top panel (Figure 10.1a) shows an example of different levels of steady-state responding under baseline and treatment conditions. The bottom panel (Figure 10.1b) shows an example of different behavioral trends under baseline and treatment conditions.

subjects in these studies complete a series of choice trials in which they decide between a smaller, immediate reward and a larger, delayed reward (e.g., "Would you prefer $100 delivered immediately or $200 delivered in one week?"). By examining subjects' choices across many different delays and reward magnitudes, researchers have determined that subjects frequently discount rewards in a pattern that can be nicely described by a hyperbola-like mathematical function (Green & Myerson, 2004; Rachlin, Raineri, & Cross, 1991). In addition, numerous studies on delay discounting have examined the extent to which different groups discount delayed rewards differently. Bickel, Odum, and Madden (1999), for instance, found that smokers discounted delayed rewards faster than nonsmokers, and Green, Fry, and Myerson (1994) observed that children discounted delayed rewards faster than young adults, who, in turn, showed greater discounting than older adults.

Although most researchers who study delay discounting present visual data that provide information on individual and group differences in discounting (for an example see Figure 10.2), many also include statistical analyses that give a quantitative description of the differences among individuals and groups (e.g., Myerson, Green, & Warusawitharana, 2001). Although some single-subject proponents have suggested that the use of statistical analysis is not necessary and may actually be detrimental when attempting to identify "real" relations among variables (e.g., Baer, 1977; Michael, 1974; Perone, 1999), others have suggested that statistical analysis can be a useful addition to the experimental toolkit of single-subject researchers (e.g., Crosbie, 1999; Davison, 1999).

METHODS OF GENERALIZATION

Since the introduction of inferential statistics by Fisher and others in the early 1900s, researchers have focused their efforts on obtaining samples of subjects that supposedly represent the greater populations from which the subjects came. Most often this strategy entails choosing a large number of subjects, collecting data, and then inferring the extent to which any observed findings are likely to be reproduced in the greater population. Thus, when using college students, for instance, researchers often try to obtain a large sample of subjects that represents in as many ways as possible the complete population of college students they are attempting to study. By doing so, large-N researchers suggest that any findings they obtain will not only describe the sample of subjects in their study, but might also describe the larger population of subjects who did not participate in the study. Moreover, by using a relatively diverse sample of subjects, large-N researchers suggest that observed outcomes might even generalize to other settings or times.

In contrast, single-subject researchers rely on replication, which ultimately serves two important functions (Johnston & Pennypacker, 1993). First, replication helps researchers determine whether their observations are replicable under similar conditions. Replicating a set of observations under similar conditions increases one's certainty regarding the "real" relations among variables. Second, single-subject researchers also want to know whether their observations will hold under different conditions, which provides information regarding the generality of a set of findings. In an attempt to answer these questions, single-subject researchers focus on two types of replication: *direct replication* and *systematic replication*.

Direct replication entails repeating the original procedures of a study as closely as possible, in hopes of determining whether an observation is reliable. As Johnston and Pennypacker (1993) noted, this type of replication may take many forms. For example, a researcher might be interested in whether a single subject continues to respond in the same way, under the same treatment condition, during a single experimental session (*within-session replication*). In contrast, a researcher might want to know whether two or more subjects in the same experiment respond similarly under the same treatment conditions

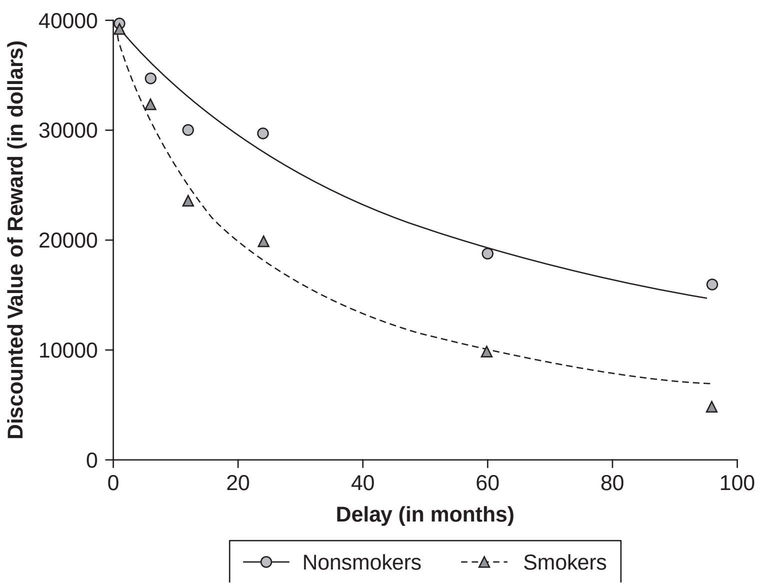

Figure 10.2 Shows hypothetical discounting data for smokers and nonsmokers. The filled circles represent median indifference points (i.e., the value at which a smaller, immediate reward is subjectively equivalent to a larger, delayed reward) for nonsmokers, and the open triangles represent median difference points for smokers. The solid and dashed lines represent the best fitting regression functions for nonsmokers and smokers, respectively.

(*within-experiment replication*). Or a researcher who earlier conducted a study with college women, for example, may want to know whether another researcher observed the same pattern of responding when replicating a study with a different sample of college women (*within-literature replication*). In each case, a researcher replicates some aspect of a study as closely as possible and notes the resulting outcomes. Finally, although direct replications focus mostly on the reliability of a set of findings, they may provide tentative information regarding generality if a series of studies occur at different times or use slightly different subjects.

Systematic replication, in contrast, occurs when researchers repeat a study but introduce conditions that are different from the original study. Remember that the primary purpose of systematic replication is to determine the generality of a specific outcome. To illustrate, consider a series of recent studies on a new instructional method known as interteaching (Boyce & Hineline, 2002). In one study, Saville, Zinn, and Elliott (2005) examined in a controlled laboratory setting whether college students in an interteaching condition received better scores on a short quiz than students in a reading, lecture, or control condition. Saville et al. observed that students in the interteaching condition answered correctly a significantly greater percentage of questions than students in the other three conditions.

To examine the generality of these results, Saville, Zinn, Neef, Van Norman, and Ferreri (2006) conducted a systematic replication in which they compared interteaching to lecture in a regular college class. In support of Saville et al.'s (2005) findings, Saville et al. (2006) also observed that students typically performed better on exams following exposure to interteaching than when they had encountered the same information via a traditional lecture. Thus, the efficacy of interteaching as a potentially effective alternative to traditional lectures seems to generalize beyond controlled laboratory settings into less controlled classroom environments.

Ultimately, by conducting both direct and systematic replications, single-subject researchers can state with confidence the extent to which the variables they study are related and whether their observations are likely to hold under different conditions. In this way, single-subject researchers systematically build a knowledge base that allows them to learn more about basic behavioral processes and apply this knowledge in hopes of modifying socially meaningful behavior.

TYPES OF SINGLE-SUBJECT DESIGNS

In general, there are two categories of single-subject designs: (a) *reversal designs*, in which a researcher introduces and subsequently removes one or more treatments; and (b) *multiple-baseline designs*, in which a researcher staggers the introduction of a treatment across behaviors, subjects, or settings. Each of these designs allows researchers to obtain information on the functional relations among variables.

Reversal Designs

A-B Design

The most basic single-subject design is the A-B design, in which researchers first measure the dependent variable under a baseline (A) condition to see how a subject behaves "normally," and subsequently under a treatment (B) condition to see if the subject's behavior changes. To illustrate, consider once again a hypothetical study in which researchers examined the effects of goal setting on the execution of various football skills (e.g., blocking; see Smith & Ward, 2006). Using an A-B design, the researchers would first measure how each football player blocked correctly in the absence of goal setting. Then, after a player's performance had stabilized, the researchers would ask the player to set goals for himself and measure any concomitant changes in blocking (see Figure 10.3). If the introduction of goal setting was accompanied by meaningful changes in blocking, the researchers might conclude that their independent variable was responsible for the changes. Unfortunately, because the treatment in A-B designs is not subsequently withdrawn (reversed back to baseline), researchers cannot determine with any certainty that changes in the dependent variable during the B phase were due to the treatment or to other extraneous factors that occurred at the same time. For this reason, researchers rarely use A-B designs if their goal is to identify functional relations.

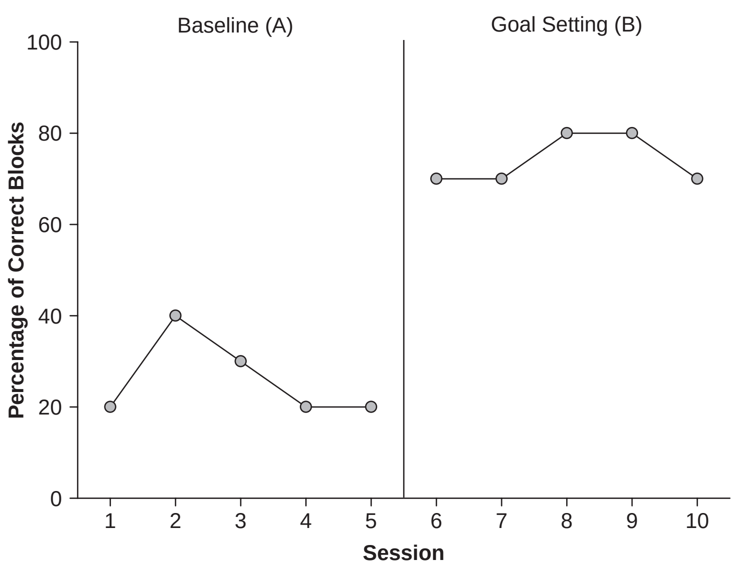

Figure 10.3 Shows an example of an A-B design.

A-B-A Design

The A-B-A, or withdrawal, design is the logical extension of the A-B design. With A-B-A designs, researchers measure the dependent variable under baseline, introduce the treatment, and then reintroduce the baseline condition again (i.e., withdraw the treatment). Thus, in the case of the previous example, researchers would measure blocking under baseline, introduce goal setting, and then subsequently remove the treatment and see how blocking performance changed (see Figure 10.4). Whereas the A-B design would not necessarily allow the researchers to draw conclusions regarding functional relationships between their variables, the A-B-A design is the simplest single-subject design that allows a researcher to make such functional statements. Specifically, if the dependent variable (e.g., proper blocking) changes *only* when the treatment (e.g., goal setting) is introduced and subsequently removed, researchers can be relatively certain that their manipulations were responsible for the observed changes.

A-B-A-B Design

An extension of the A-B-A design is the A-B-A-B design, in which a treatment is introduced after an initial baseline condition, subsequently removed, and then introduced again (see Figure 10.5). Thus, whereas the A-B-A design ends with baseline (i.e., the absence of a treatment), the A-B-A-B design ends with behavior occurring under a treatment condition. The addition of a second treatment condition may be important for at least two reasons. First, if the dependent variable changes once again with the reintroduction of the second B phase, a researcher can be even more certain that the treatment was responsible for the observed changes. Second, the reintroduction of a treatment condition may be important for ethical reasons. Consider, for instance, a subject who is engaging in severe SIB. If a treatment is effective at reducing the incidence of SIB, ending a study in a baseline phase might put the subject at risk if the SIB returns. Of course, depending on the purpose of the study, researchers might extend the A-B-A-B design to include more treatment presentations and withdrawals (e.g., A-B-A-B-A-B design).

Multiple-Treatment Designs

In contrast to the preceding designs in which researchers introduce only one treatment condition, multiple-treatment designs are those in which researchers introduce more than one treatment condition. Examples of multiple-treatment designs would include A-B-A-C and A-B-A-C-A-D designs, in which a researcher alternates several treatments with baseline; A-B-C-D-B-C-D designs, in which a researcher alternates several treatments without reestablishing baseline before and after each; and A-B-A-C-A-BC-A designs, in which a researcher examines the effects of more than one treatment, individually and then together (see Figure 10.6 for an example of an A-B-A-C design). Again, depending on the purpose of the study, researchers may choose to use variations of these designs.

When considering which multiple-treatment design to use, single-subject researchers often have to consider the pros and cons of using each. For example, with A-B-A-C

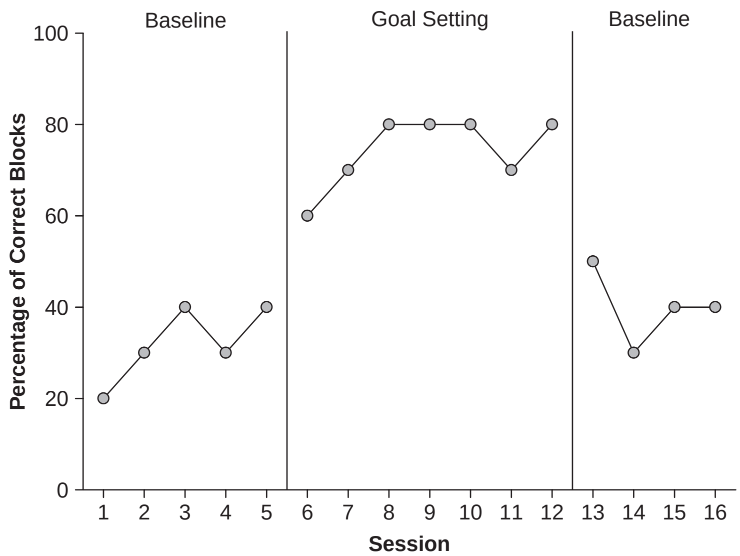

Figure 10.4 Shows an example of an A-B-A design.

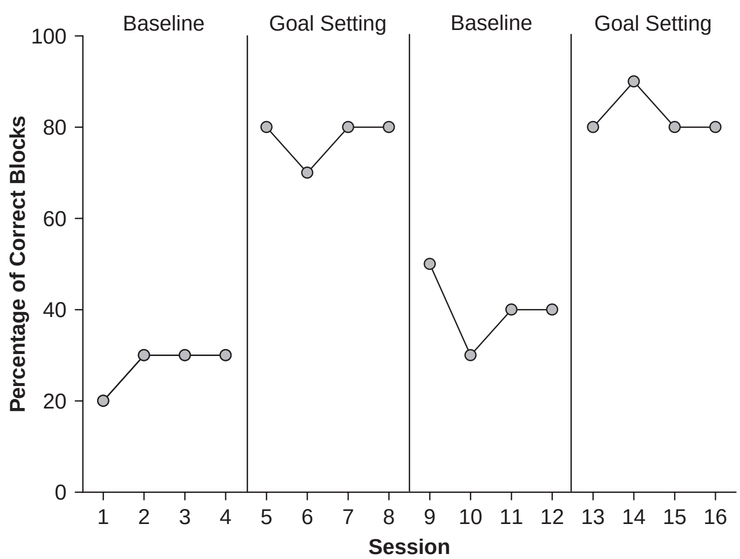

Figure 10.5 Shows an example of an A-B-A-B design.

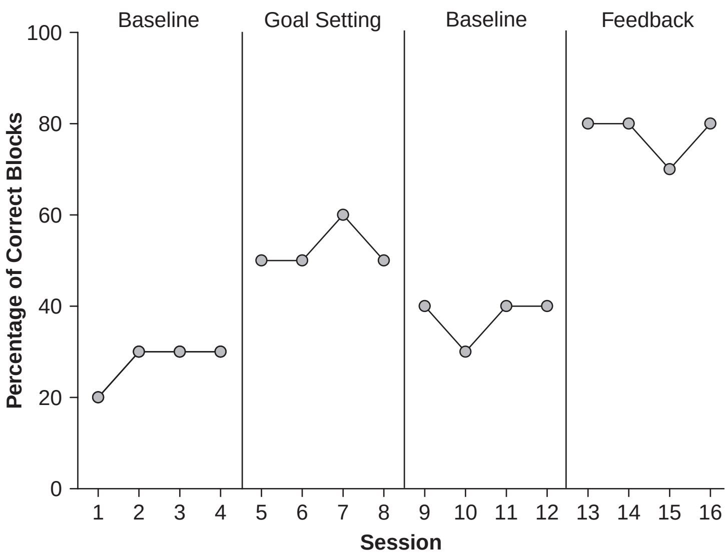

Figure 10.6 Shows an example of an A-B-A-C design.

and A-B-A-C-A-D designs, researchers may be able to determine with more certainty the extent to which each treatment affected their dependent variables. By reestablishing baseline before and after each treatment condition, researchers are more likely to see the "real" effects of each treatment and not the effects of carryover from a preceding condition. However, reestablishing baseline after each treatment condition may be time-consuming, which could be an issue depending on the resources researchers have at their disposal (e.g., Cooper et al., 1987). Similarly, with an A-B-C-D design, for example, researchers may be able to study multiple treatments in less time, but carryover or order effects may make it difficult to determine how each treatment in isolation affects the dependent variable. Fortunately, counterbalancing the order of conditions with all reversal designs (e.g., using an A-B-A-C design with some subjects and an A-C-A-B with others) may help researchers determine how different treatments are affecting the behavior of interest. Ultimately, the type of multiple-treatment design that researchers choose to use in their studies may depend on the specific purpose of their study.

Multiple-Baseline Designs

Multiple-baseline designs offer an alternative to reversal designs when researchers (a) are studying behavior that is irreversible or (b) do not wish to remove a treatment but still wish to identify functional relations between their variables (Baer et al., 1968; Cooper et al., 1987). In essence, multiple-baseline designs entail the measurement of behavior under more than one baseline condition and the staggered introduction of a treatment to see if behavior changes *only* when the treatment is introduced.

For example, consider a study in which a researcher wanted to know whether some treatment decreased SIB in three developmentally disabled adults. Because a reversal design would entail the removal of a potentially effective treatment, which, as I discussed earlier, may carry ethical considerations, a researcher could alternatively use a multiple-baseline design. First, the researcher would measure the "normal" rate of SIB for each subject in the absence of treatment. Once the rate of SIB for each subject had stabilized, the researcher would introduce the treatment for one subject but not the others. Then, at a later point, the researcher would introduce the treatment for a second subject, while the third subject continued under baseline. Finally, the researcher would introduce the treatment for the last subject. If each subject's behavior changed when—*and only when*—the treatment was introduced, the researcher could be relatively certain that the treatment, and not some other factor, was responsible for the observed changes (see Figure 10.7).

There are three general variations of the multiple-baseline design (see Cooper et al., 1987). The preceding example constitutes a *multiple baseline across subjects design*, in which the same behavior is investigated in different subjects in the same setting. A *multiple baseline across settings design* examines the same behavior in the same subject, but in different settings. Finally, a *multiple baseline across behaviors design* examines different behaviors in the same subject and setting. Each of these designs provides useful information regarding functional relations as well as information regarding the generality of a particular treatment. Finally, although multiple-baseline designs do not allow researchers to examine the effects of an independent variable directly (i.e., by introducing and removing a treatment) and, thus, are considered by some to be weaker than reversal designs for examining functional relations (Cooper et al., 1987), these designs still provide a useful and practical way for researchers and practitioners to examine the extent to which their treatments produce changes in a behavior of interest.

SUMMARY

The history of psychology is replete with studies in which early, influential researchers examined interesting phenomena using a small number of subjects under tightly controlled experimental conditions. However, the introduction of well-known statistical procedures in the early 20th century led to the emergence of large-N, group-oriented designs, which have exerted a stronghold on psychological research for nearly a century. Nevertheless, single-subject designs continue to provide a viable alternative for researchers wishing to study functional relations and learn more about interesting psychological phenomena. Ultimately, though, which design researchers choose to use in their studies may depend on the specific experimental questions they ask.

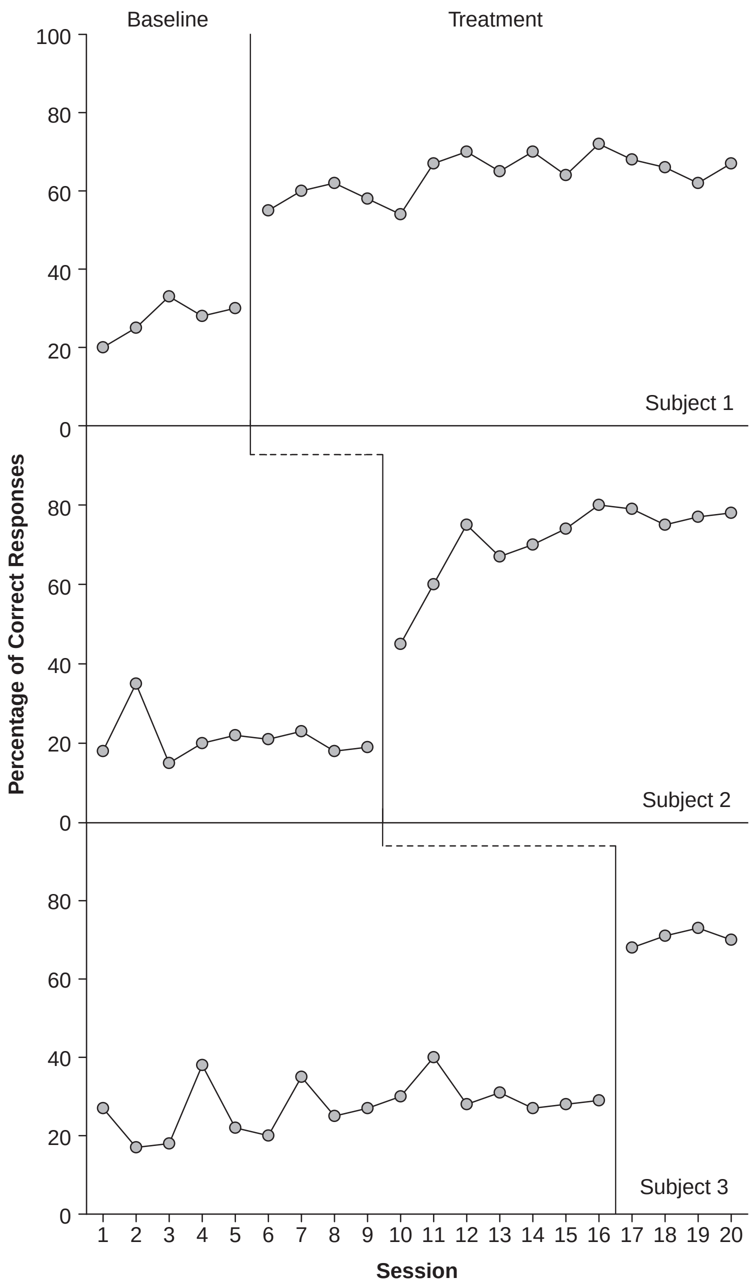

Figure 10.7 Shows an example of an A-B-A-C design.

REFERENCES AND FURTHER READINGS

Baer, D. M. (1977). Perhaps it would be better not to know everything. *Journal of Applied Behavior Analysis, 10,* 167–172.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). Boston: Allyn & Bacon.

Baron, A., & Perone, M. (1998). Experimental design and analysis in the laboratory study of human operant behavior. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 45–91). New York: Plenum Press.

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology, 146,* 447–454.

Blampied, N. M. (1999). A legacy neglected: Restating the case for single-case research in cognitive-behaviour therapy. *Behaviour Change, 16,* 89–104.

Boring, E. G. (1954). The nature and history of experimental control. *American Journal of Psychology, 67,* 573–589.

Boring, E. G. (1957). *A history of experimental psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Boyce, T. E., & Hineline, P. N. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangements in the college classroom. *The Behavior Analyst, 25,* 215–226.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill.

Crosbie, J. (1999). Statistical inference in behavior analysis: Useful friend. *The Behavior Analyst, 22,* 105–108.

Davison, M. (1999). Statistical inference in behavior analysis: Having my cake and eating it? *The Behavior Analyst, 22,* 99–103.

Fisher, R. A. (1925). *Statistical methods for research workers.* Edinburgh, UK: Oliver & Boyd.

Goodwin, C. J. (2005). *A history of modern psychology* (2nd ed.). Hoboken, NJ: Wiley.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130,* 769–792.

Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A lifespan comparison. *Psychological Science, 5,* 33–36.

Hall, R. V., Lund, D., & Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis, 1,* 1–12.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Orlando, FL: Harcourt.

Kimble, G. A. (1999). Functional behaviorism: A plan for unity in psychology. *American Psychologist, 54,* 981–988.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7,* 560–572.

Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis, 7,* 647–653.

Miller, R. L. (2003). Ethical issues in psychological research with human participants. In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology* (pp. 127–150). Malden, MA: Blackwell.

Mitchell, M. L., & Jolley, J. M. (2004). *Research design explained*. Belmont, CA: Thomson Wadsworth.

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior, 76,* 235–243.

Perone, M. (1999). Statistical inference in behavior analysis: Experimental control is better. *The Behavior Analysis, 22,* 109–116.

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior, 55,* 233–244.

Robinson, P. W., & Foster, D. F. (1979). *Experimental psychology: A small-N approach*. New York: Harper & Row.

Saville, B. K., & Buskist, W. (2003). Traditional idiographic approaches: Small-N research designs. In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology* (pp. 66–82). Malden, MA: Blackwell.

Saville, B. K., Zinn, T. E., & Elliott, M. P. (2005). Interteaching vs. traditional methods of instruction: A preliminary analysis. *Teaching of Psychology, 32,* 161–163.

Saville, B. K., Zinn, T. E., Neef, N. A., Van Norman, R., & Ferreri, S. J. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39,* 49–61.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow database of social psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515–530.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Boston: Authors Cooperative.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1966). Operant behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 12–32). Englewood Cliffs, NJ: Prentice Hall.

Smith, S., & Ward, P. (2006). Behavioral interventions to improve performance in collegiate football. *Journal of Applied Behavior Analysis, 35,* 385–391.

Staats, A. W. (1999). Unifying psychology requires new infrastructure, theory, methods, and research agenda. *Review of General Psychology, 3,* 3–13.

Stanovich, K. E. (2001). *How to think straight about psychology* (6th ed.). Boston: Allyn & Bacon.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797–811.

Sternberg, R. J. (2005). *Unity in psychology: Possibility or pipedream?* Washington, DC: American Psychological Association.

11

QUALITATIVE RESEARCH

JENNIFER FEATHERSTON
University of Arkansas

Qualitative research provides rich and detailed accounts of matters under examination in ways that other, more traditional types of research are unable to express. One of the most important distinctions that sets qualitative psychological research apart from more traditional types of research is that qualitative research is holistic, in that researchers study phenomena in their entirety rather than narrowing the focus to specific defined variables. In this chapter, you will be presented with information that beginning students of qualitative research and prospective researchers should know.

Qualitative research is exciting and vital for producing new information. It is a highly satisfying endeavor in that it connects us with individuals, activities, and situations that are of importance to research. It allows for the careful exploration of dimensions in the social world that can give the research a depth of information not always obtainable through traditional quantitative research. Qualitative research shows us pictures of everyday life including the understandings, ideas, and experiences of the participants that are rich in significance and provide in-depth understanding. These elements are directly factored into the researcher's analyses and explanations.

Qualitative analyses are nonstatistical methods using inductive logic. Qualitative research develops theory whereas quantitative research develops and tests theory. Qualitative research describes meaning or discovery whereas quantitative establishes relationship or causation. In qualitative research the researcher is explicitly a part of the data-gathering process whereas in quantitative, the researcher is formally an independent entity. Qualitative research uses communication and observation whereas quantitative uses instruments. Qualitative research uses unstructured data collection whereas quantitative uses structured data collection.

There are many reasons that a qualitative study may be selected over a quantitative study. Some people may simply prefer qualitative research, thinking it is suited to their personal style. Others feel that to make real changes, qualitative is more appropriate. Some people state that they are not comfortable with statistics or math and that they are stronger with their verbal and conceptual skills. Others want to use their research to bring about social, political, and economic change and believe that by engaging participants in the process, they can more easily achieve this change. Inviting cooperation by participants is almost always forbidden in most quantitative research. Although there is validity in some of these reasons, others are simply based on misinformation or lack of knowledge.

HISTORY OF QUALITATIVE RESEARCH

Qualitative study has its roots in anthropology and sociology; the first researchers were anthropologists who conducted ethnographies of primitive cultures. These in-depth, lengthy studies occurred in the native tribes' homeland and the researcher was as noninvasive as possible, often living among the people or studying them from a distance, so as not to influence the everyday activities of the group.

Franz Boas, who was born in Germany in 1858, is called "The Father of American Anthropology" because of his contributions to the discipline after he moved to America. In 1883 Boas went to Baffin Island (off the northern coast of Canada) to conduct geographic research on the impact of the physical environment on native Inuit migrations. Boas lived and worked closely with the Inuit peoples and developed an interest in the way the people lived. In a diary entry dated December 23, 1883, Boas wrote,

> I often ask myself what advantages our "good society" possesses over that of the "savages" and find, the more I see of their customs, that we have no right to look down upon them.…We have no right to blame them for their forms and superstitions which may seem ridiculous to us. We "highly educated people" are much worse, relatively speaking.…

Boas' studies were in contrast to the other interactions with "primitive" peoples by his contemporaries, many of whom were missionaries, explorers, and colonial bureaucrats, who had intentions of influencing or changing the ways these people lived. Boas was a cultural relativist who believed the object of anthropological study is to describe the knowledge that members use to make sense within their own culture.

Qualitative research evolved through a series of historical "moments" (Denzin & Lincoln, 1994) during the period from 1900 to World War II called the *Traditional Period*. During this period, anthropologists like Margaret Mead, Bronisław Malinowski, and Alfred Radcliffe-Brown spent prolonged periods of time living among natives and compiling extensive ethnographies (Rosaldo, 1989). They developed fieldwork practices that are still used by qualitative researchers today, including *participant observation*, *interviewing*, and *artifact gathering*, which are described in more detail in this chapter.

During this same time, sociologists were also exploring the integration of qualitative research methods in their field. At the University of Chicago, sociologists began using their city as a social laboratory to study the everyday lives of Chicagoans (Vidich & Lyman, 1994). Termed *Chicago Sociology*, the research engaged sociologists for three decades, during which time they produced many urban ethnographies capturing life in the city. The studies utilized techniques similar to those used by the anthropologist studying in faraway lands; only the people and the settings varied.

The period from World War II until the mid-1970s is called the *Modernist Phase* (Denzin & Lincoln, 1994); it was during this time that researchers formalized qualitative research methods. Attempts were made to establish a fit between expectations for validity, reliability, and generalizability and models of doing research. Several revolutionary books were published including Glaser and Strauss' *The Discovery of Grounded Theory* (1967) and Blumer's *Symbolic Interactionism* (1969).

The period from the mid-1970s into the early 1980s is called *The Moment of Blurred Genres* (Denzin and Lincoln, 1994) because of the wide range of paradigms, methods, and strategies from different disciplines. During this time of great debate between qualitative and quantitative camps, qualitative research continued to make its mark as a valid form of scientific research. Journals began publishing more qualitative studies, and qualitative researchers were invited to present at national conferences, thus increasing the legitimacy of the research.

The stage spanning the mid-1980s until the early 1990s was called the *Crisis of Representation* (Denzin & Lincoln, 1994) because of the difficulty in representing truth through text produced from field notes and interviews. The argument is that "understandings of human experience are always processed through language, and language is inherently unstable" (Denzin, 1989). Although each of these historical periods has come and passed, the methods developed continue to be used by researchers today.

THEORY

Hatch (2002) recommends that beginning researchers develop their own definitions of what they believe qualitative research to be in an attempt to establish "conceptual boundaries" for their work. Definitions range from descriptive formulations to product-oriented statements, but the form is not as important as the content. A comprehensive definition, devised by Bogdan and Taylor (1975), states that qualitative research "directs itself at settings and the individuals within those settings holistically; that is, the subject of the study, be it an organization or an individual, is not reduced to an isolated variable or to an hypothesis, but it is viewed instead as part of a whole" (p. 2).

Attributes of Qualitative Research

Many attributes characterize qualitative research. The following discussion introduces the beginning researcher to some of the qualities that distinguish qualitative research from other types of inquiry: centrality of meaning, emergent design, extended firsthand engagement, inductive data analysis, natural settings, participant perspectives, reflexivity, researcher as a data-gathering instrument, subjectivity, and wholeness and complexity.

Centrality of Meaning

Blumer's (1969) symbolic interactionist theory is a way to systematically explore understandings. Three premises of symbolic interactionism underlie centrality of meaning: (a) humans react to things based on the meanings the things hold, (b) the meanings come from social interactions with other people, and (c) these meanings can change as people encounter new people and things. Although not all qualitative research uses the symbolic interactionist framework, it is a useful way for researchers to understand the way individuals construct meanings in order to participate in society.

Emergent Design

The nature of qualitative research makes it almost impossible to completely design an experiment in advance. Once the researcher enters the environment that is being studied, research designs may change and new designs may emerge. Although researchers may disagree on the extent to which research proposals should be followed, most agree that elements of the design such as research questions and methods may change as the research progresses.

Extended Firsthand Engagement

Extended firsthand engagement refers to spending extensive time in the field researching. Hatch (2002) argues that if researchers are to understand participant perspectives in natural contexts, they must spend enough time in the environment to feel certain they are reporting what they claim. Although qualitative inquiry appeals to a great number of people, few are willing to spend the time needed to produce first-rate research.

Inductive Data Analysis

Qualitative researchers first collect as much information as they can from the setting they are studying and then look for patterns and relations in what they have amassed. It is similar to putting a puzzle together without the advantage of looking at a completed picture of what it should look like when finished. The picture becomes evident as the pieces are placed together. However, qualitative data analysis also involves a deductive element because as the patterns are revealed in the data, hypothetical categories are created, and the data are then read deductively to see if these categories are supported by the original data set (Erikson, 1986).

Natural Settings

The objects of study in qualitative research are the lived experiences of real people in real settings. Therefore, researchers collect information on how individuals make sense of their lives. In traditional research, the examiner has control over the environment and sets up the experiment in a way that controls for the variables of interest. Conversely, qualitative researchers aim to explore human behavior within the most natural setting.

Participant Perspectives

Because qualitative researchers want to understand the world from the perspectives of the people who live in it, they look at the ways people respond to their environments. Researchers are interested in what is happening in the environment, how the participants are responding, and what the interaction between person and environment means to the participant. Researchers should make an effort to highlight the voices of the participants in any good qualitative report.

Reflexivity

In traditional research, it is expected that the researcher remain the "objective scientist." However, in qualitative research it is impossible to play this role; the researcher is often part of the world of study. Reflexivity is the "process of personally and academically reflecting on lived experiences in ways that reveal deep connections between the writer and his or her subject" (Goodall, 2000).

Researcher as a Data-Gathering Instrument

The traditional role for a quantitative researcher is to be nonexistent. It would be ideal if a participant acted exactly as he or she would if you weren't present. The problem is that this tends to ignore any differences made in the environment by the researcher. Qualitative research says document such differences and explicate them. It is difficult to get participants' views without interacting. However, the primary data in qualitative research is often gathered by the researcher. Thus, the researcher is considered the data-gathering instrument, and acts in ways similar to questionnaires, checklists, scales, tests, and other measuring devices. Data may be collected directly from participants in the form of field notes from participant observation or notes from transcripts or interviews. Information can also be collected unobtrusively via artifacts from the research environment or records related to the social experience under investigation. Human competences necessary to take part in living are the same as those that allow qualitative researchers to make sense of the actions, intentions, and understandings of the participants they are examining. Those reading the report need to know about the instrument, so the researcher should describe relevant aspects of self, biases and assumptions, expectations, and relevant history. The researcher may want to keep track of personal reactions and insights into self and post them in a separate journal.

Subjectivity

Qualitative researchers must rely on subjective interpretation because the inner states of their participants are often not directly observable. They would argue that complete objectivity is not possible in any scientific study but that all findings generated from the research are grounded in empirical evidence from their data. The goal is to apply subjectivities in ways that make it possible to understand the unstated motives and assumptions of the participants.

Wholeness and Complexity

Qualitative methods allow researchers to examine unique, dynamic, and complex social settings without reducing them to small pieces of fragmented variables. Qualitative data are "objects, pictures, or detailed descriptions that cannot be reduced to numbers without distorting the essence of the social meanings they represent" (Hatch, 2002, p. 9). The consumer of the research is able, through the detailed data, to be transported into the social situation under examination.

Research Paradigms

Research paradigms are fundamentally different belief systems about how the world is ordered, what we may know about it, and how we may know it. We examine five research paradigms according to ontology, epistemology, methodology, and products. The intention is to give novice researchers a guide for exploring assumptions about what research is and how it works. For a more in-depth discussion of these research paradigms, see Denzin and Lincoln (1994) and Guba and Lincoln (1994).

Positivist Paradigm

The positivist (or realist) believes in an objective universe that has order regardless of human perceptions. Reality is out there to be studied, captured, and understood. Positivists believe that the world has order and that it is possible, through research, to discover that order. Researchers gain knowledge through experiments, quasi experiments, surveys, and correlational studies. The products, or forms of knowledge produced, include facts, theories, laws, and predictions.

Postpositivist Paradigm

In the postpositivist paradigm (or critical realism), reality exists but is never fully apprehended, only approximated. The researcher knows an approximation of reality, as the researcher is the data-collection instrument. Knowledge is gained through rigorously defined qualitative methods, frequency counts, and low-level statistics. Forms of knowledge produced include generalizations, descriptions, patterns, and grounded theory.

Constructivist Paradigm

According to the constructivist paradigm, multiple realities are constructed. Knowledge is a human construction, and the researcher and participant(s) coconstruct understandings. Knowledge is gained through naturalistic qualitative methods that result in case studies, narratives, interpretations, and reconstructions.

Critical/Feminist Paradigm

Those who subscribe to the critical/feminist paradigm believe that the apprehended world makes a material difference in terms of race, sex, and class. According to this view, knowledge is subjective and political and researchers' values frame inquiry. Knowledge is gained through transformative inquiry that yields value-mediated critiques that challenge existing power structures and promote resistance.

Poststructuralist Paradigm

In the poststructuralist paradigm, order is created within individual minds to ascribe meaning to a meaningless universe. Because (according to this perspective) there is no "truth" to be known, researchers examine the world through textural representations of it. Knowledge about the world is gained through methods like deconstruction, genealogy, and data-based, multivoiced studies. These methods result in reports in the form of deconstructions, genealogies, and reflexive, polyvocal texts.

The five paradigms introduced here are ways of thinking about the order of the world, what is deemed knowledge, and how that knowledge can be expanded. It is important to note that after a researcher has selected the paradigm that makes sense to him or her, the positions taken by other researchers working within other paradigms will not make sense. Starting with the positivist paradigm and moving toward the poststructuralist paradigm, the paradigms move from modern to post-modern. The modernist paradigms are more quantitative in nature and the postmodernist paradigms are more qualitative. Hatch (2002) holds that most students and most individuals in Western society would identify with the positivist paradigm, but beginning researchers are encouraged to explore their own beliefs about how the world is ordered and how it can be known.

METHODS

Types of Qualitative Inquiry

There are many kinds of qualitative inquiry, and each type has its own unique characteristics. This section includes descriptions of several types of qualitative research from which the beginning researcher may select when choosing to do a qualitative study.

Naturalistic Inquiries

Simply stated, naturalistic methods mean the same as qualitative research. The goal is to capture natural activity in natural settings. Naturalistic inquiry is a specific type of naturalist (or qualitative) method. When designing a naturalistic study, Lincoln and Guba (1985) recommend that researchers do the following: (a) determine the focus of the inquiry, (b) determine whether the focus and the paradigm are an appropriate match, (c) determine the fit of the inquiry paradigm to the theory selected to guide the study, (d) determine where the data will come from, (e) determine the steps of the inquiry, (f) determine what instrument(s) will be used to collect the data, (g) plan data collection, (h) plan data analysis techniques, (i) plan the logistics, and (j) plan the trustworthiness of the study. Naturalistic inquiry is the archetype for constructivist qualitative research.

Ethnographies

Ethnographic inquiries seek to describe culture or parts of culture from the point of view of cultural insiders. Classic anthropological studies, like Boas' examination of the Inuit people, are ethnographies. Researchers collect data through participant observation, interviewing, and artifact collection. The purpose of ethnography is to account for the behavior of people by describing what they know and how that knowledge enables them to behave appropriately given dictates of common sense in their community.

Macroethnographies and microethnographies are subcategories of ethnography. Macroethnography is merely another word for ethnography (described previously), but microethnographies are not (as the name suggests) ethnographies on a smaller scale. They are most often carried out by sociolinguists and other researchers interested in verbal and nonverbal communication. They collect data through videotaped face-to-face exchanges that reveal the linguistic rules that participants use to construct meanings (see Bremme & Erikson, 1977). These studies are frequently based within the positivist paradigm.

Ethnomethodology

Ethnomethodology focuses on everyday life and is the study of norms, understanding, and assumptions that are usually taken for granted. There is an effort to enrich or extend the study of a culture and its people through personal experience, direct participation, and immersion, thereby achieving an emic (insider's) perspective on the people and the culture. Interactions are reported autobiographically. Ethnomethodologies function within the postpositivist paradigm and produce theories to explain contextualized human behavior. An example of an ethnomethodological inquiry is an examination of restoring order to a classroom after a disruption has taken place.

Symbolic Interaction

Symbolic interactionist studies examine the common symbols that give meaning to human interaction within a social group context. These inquiries consist of exploration and inspection elements. The exploration component includes collecting observations, interviews, life histories, letters, diaries, public records, and group discussions. In the inspection phase, the researcher discriminates analytic elements and isolating relations between elements (Blumer, 1969). Of the different types of qualitative research, symbolic interaction stands out because it represents a break from positivism.

Hermeneutics

Hermeneutics is the reinterpretation of stories, texts, and narratives so as to obscure their original purpose and intent in a particular situational context. Researchers examine the social and cultural conditions or circumstances in which the text was created. Methods used to gather experiential experiences from participants include protocol writing, interviewing, observing, and studying descriptions in literature and the arts. Protocol writing simply consists of asking a participant to write down his or her experiences, and interviewing is done by gathering narrative material through conversation. To observe is to collect anecdotes of experience through close observation. Researchers can look to literature and the arts for rich insight into the nature of the phenomenon under investigation. Extensive archival data exist in poetry, novels, plays, biographies, works of art, diaries, journals, and logs and provide insight into the culture as well as individual experiences. Hermeneutic phenomenology is a constructivist approach because it assumes that although many collectively constructed realities exist at the same time, individuals' personal experiences ought to be the object of study.

Grounded Theory

Grounded theory was developed by Glaser and Strauss (1967) and provides a detailed framework for collecting and analyzing qualitative data in methodical ways. It is a process of generating theory rather than a particular theoretical context. Grounded theory requires the use of inductive derivation and the development of explanatory theory from a documented observational date. This type of inquiry involves a researcher looking at documents (such as field notes) for indicators of categories in events and behavior. The researcher then names and codes them and compares codes to find consistencies and differences. Consistencies between codes (similar meanings or pointing to a basic idea) reveal categories. Although early researchers in this field used to cut apart their notes and make piles of different categories, there are now computer programs that do the work. Grounded theory is a postpositivist method because it assumes that thorough methods can be used to discover estimates of reality that are empirically represented in carefully collected data.

Participant Observation Studies

Although participant observation studies are similar to ethnographies, they do not require the researcher to spend as much time in the field as ethnographic studies. They are also narrower in focus, and the researcher enters the field with specific interests and questions. Participant observation requires that the researcher become a participant in the culture or context being observed. The literature on participant observation discusses how to enter the context, the role of the researcher as a participant, the collection and storage of field notes, and the analysis of field data. Participant observation can require many days or months of intensive work because the researcher needs to become accepted as a natural part of the culture in order to assure that the observations are of the natural phenomenon. Participant observation fieldwork includes data collection strategies that researchers can use with any of the qualitative research paradigms.

Interview Studies

Although interviewing can be a component of many participant observation studies, it also can be used as the primary data-collection strategy in a qualitative project. In-depth interviews include both individual interviews (e.g., one-on-one) as well as “group” interviews (including focus groups). The data can be recorded in a wide variety of ways including stenography, audio recording, video recording, or written notes. In-depth interviews differ from direct observation primarily in the nature of the interaction. In interviews it is assumed that there is a questioner and one or more interviewees. The purpose of the interview is to probe the ideas of the interviewees about the phenomenon of interest. Unstructured interviewing involves direct interaction between the researcher and a respondent or group. It differs from traditional structured interviewing in several important ways. First, although the researcher may have some initial guiding questions or core concepts to ask about, there is no formal structured instrument or protocol. Second, the interviewer is free to move the conversation in any direction of interest that may come up. Consequently, unstructured interviewing is particularly useful for exploring a topic broadly. However, there is a price for this lack of structure. Because each interview tends to be unique with no predetermined set of questions for all respondents, it is usually more difficult to analyze unstructured interview data, especially when synthesizing across respondents. Qualitative interview studies can be undertaken from all but the positivist paradigm.

Focus Group Studies

Focus groups are sets of individuals with similar characteristics or shared experiences who sit down with a moderator to discuss a topic. The discussion is particularly effective in providing information about the way people think or why they feel the way they do. The focus group is a special type of group in terms of purpose, size, composition, and procedures. The group is typically composed of seven to ten participants who are unfamiliar with each other. The primary reason focus groups are successful is that they tap into human tendencies (Krueger, 1988).

The people included in the focus group possess certain characteristics, provide data of a qualitative nature, and are led in a focused discussion. Focus groups serve many purposes: generating information for questionnaires, producing a needs assessment, testing new programs, discovering what people consider when making decisions, establishing community standards, evaluating existing programs, understanding an organization’s image, getting a response from a mail survey, assessing a product, or providing feedback to administrators.

To make the focus group experience more appealing to prospective members, inducements such as prestige, responsibility, curiosity, and cash are often used to elicit participation. The moderator of the group is there to lead the discussion and maintain order. Typically, the key questions will be limited to about six to ten questions, with possible subpoints within each question. The primary reason for the questions is to maintain a logical sequence to the discussion. The discussion should be limited to an hour and a half. The session should be recorded, and someone should take notes. Krueger (1998, pp. 89–90) presents a very good checklist for focus group interviews. Like individual interviews, focus group interviews can be utilized with any of the qualitative research paradigms.

Narrative Studies

Narrative study is a broad category for research that examines gathering and interpreting the stories that people use to describe their lives. Specific methods include life histories, life history research, biography, personal experience methods, oral history, and narrative inquiry. Closely tied to this is the notion of narrative analysis, which is the study of an individual’s speech. Narrative inquiry overlaps with other approaches. Where discourse analysis looks at interaction between people, narrative analysis looks at the individual. The reasoning is that *what* someone chooses to tell frames *how* he or she will be perceived. People always compare ideas about themselves and tend to avoid revealing negatives about themselves. Narrative analysis could also involve study of literature, diaries, or folklore. Because the emphasis is on the meanings individuals generate through stories, this type of inquiry best fits within the constructivist or critical/feminist paradigms.

Artifact Analysis

Although many qualitative studies incorporate artifact analysis, it is rare for artifacts other than text-based materials to be used as the primary data source. Text-based materials refer to existing documents (as opposed to transcripts of interviews conducted for the research). These can include newspapers, magazines, books, Web sites, memos, transcripts of conversations, and annual reports. Usually written documents are analyzed with some form of content analysis. Artifact analyses fit most appropriately within the postpositivist paradigm because of the static nature of the data and the reliance on the researcher as the data-collection instrument.

One example of artifact analysis is a study of radio station listening preferences. Rather than conducting an obtrusive survey or interview about favorite radio stations, the researchers went to local auto dealers and garages and checked all cars that were being serviced to see what stations the radios were currently tuned to. In a similar manner, if a researcher was interested in magazine preferences, he or she might rummage through the trash of your sample or even stage a door-to-door magazine recycling effort. Researchers need to be very careful about the ethics of this type of measurement. Because researchers may be collecting information without the respondent's knowledge, they may be violating the respondent's right to privacy.

Historical Studies

Historical Studies (or Historiographies) involve the collection and analysis of data for the purpose of reconstructing events that happened in the past. Sources for the data are classified as either primary or secondary sources. Primary sources include testimonies (oral or written), original documents, photographs, diaries, journals, drawings, and mementos. Some examples of secondary sources are elements created by others related to the event or events of relevance such as textbooks, journal articles, newspaper accounts, and public records. The researcher is responsible for authenticating the sources and deciding whether to include them in the study. Researchers who use this type of inquiry can operate within several research paradigms. Within the postpositivist paradigm, for example, researchers might generate careful descriptions of events or characters from the past.

Case Studies

Case studies are not distinct from ethnographic or participant observation studies. The definition makes the distinction that case studies are examinations of contemporary figures (as opposed to historical). A case study is an intensive study of a specific individual or specific context. For example, Freud produced case studies of several individuals as the basis for his theory of psychoanalysis. There is no single way to conduct a case study, and a combination of methods such as unstructured interviewing and direct observation can be used. Case studies fit within several qualitative paradigms, but most researchers advocate postpositivist approaches to case study research.

Action Research Projects

Action-focused research (also known as orientational) is ideological, critical, and emancipating because it is concerned with activity and change. This type of research is a practical tool for solving problems and is undertaken to investigate practices and design changes. Researchers may ask how injustice and subjugation shape peoples' experiences and understanding of the world. The goal is not to just study and understand, but to critique and change something. Researchers use change-oriented engagement strategies, such as an ideological "lens" to explicate, analyze, critique, or demystify a text or social context. Because the desired result is change, and because the researcher's values are known, this type of inquiry fits most neatly within the critical/feminist paradigm.

Sampling Strategies

Quantitative studies look at sampling in a probabilistic manner. Researchers try to get a representative sample so the results will generalize to the whole population. Qualitative researchers use sampling in this sense, but the selection process focuses on defining the variety of participants to be studied. There are several ways to choose whom to study, and the process is dynamic and ongoing; the choices of whom to study next are not in the initial plan, but are products of what is found.

Some widely used approaches in sampling include maximum variation, snowball approach or networking, extreme or deviant case, typical case, unique case, ideal case, and negative case. In maximum variation sampling, relevant dimensions vary widely within a group. There are central themes that cut across a great deal of participant or program variation, and there are shared patterns across variations. The researcher will clearly see extremes and keep them separate to avoid averaging together extremes so results represent the "average." When using the snowball approach or networking, each person studied is chosen by a previous participant. Thus researchers will see linkages between people. The researcher may ask, "Who knows a lot about__________?" or "To whom should I talk about _________?"

Using extreme or deviant case sampling allows researchers to study one or more individuals at some extreme. Researchers using this method are not interested in studying the average or the opposite. They want to study people or situations that are unusual or special in some way. Freud's examination of hysteric women is an example of this type of sampling. In this type of specialized study, it is not uncommon to have only one participant.

Researchers using typical case sampling decide what characterizes "typical" and then search out that individual. Typical case sampling seeks out the most frequently occurring situation or condition. When researchers stumble upon very rare combinations of things, they have what is called a *unique case*. *Ideal case* is simply the perfect situation.

When a researcher is looking for an exception to the emerging rule or hypothesis, she is looking for a *negative case*. This is often used in analytic induction approach when the goal is to refine generalizations by setting out to find when and where it doesn't hold true.

Although other types of sampling are used, the goal is not to know them all, but to see that there is a wide variety of possibilities. Researchers should be open to inventive ways of choosing the best ways to study participants in different situations. Sampling methods that emerge during study frequently provide the researcher with what is needed and may be better than using initial selection. However, it is very important to articulate in writing how and why a certain sampling technique was used. Researchers should provide detail as well as a strong rationale for choices made.

QUALITATIVE DATA ANALYSIS

Qualitative Validity and Reliability

In quantitative research, validity is defined as the extent to which a measuring instrument is measuring what is intended. In qualitative inquiry, many researchers reject this framework because they believe it is based on the realist assumption that there is a reality external to our perception of it. Consequently, it doesn't make sense to be concerned with whether an observation is true or false with respect to an external reality. Another way to look at this concept is to decide whether your study is trustworthy. How can a researcher persuade his or her audience that the findings of an inquiry are worth paying attention to?

Guba and Lincoln (1994) proposed four criteria for judging the trustworthiness of qualitative research and offered credibility, transferability, dependability, and confirmability as alternatives to more traditional quantitatively oriented criteria. These criteria better reflect the underlying assumptions involved in much qualitative research.

Credibility

How credible are the particular findings of the study? The goal is to demonstrate that the inquiry was conducted in such a manner as to ensure that the subject was accurately identified and described. This involves establishing that the results of qualitative research are believable from the perspective of the participant in the research. Because, from this perspective, the purpose of qualitative research is to describe or understand the phenomena of interest from the participant's eyes, the participants are the only ones who can legitimately judge the credibility of the results.

Transferability

How transferable and applicable are these findings to another setting or group of people? The purpose is to demonstrate that the applicability of one set of findings to another context rests more with the investigator who would make that transfer than with the original investigator. This is also known as *generalizability*. The qualitative researcher can enhance transferability by doing a thorough job of explaining the context of the research and the specific assumptions that were made. The person who wishes to generalize the results to a different context is then responsible for making the judgment of how sensible that is.

Dependability

How can we be reasonably sure that the findings would be replicated if the study were conducted with the same participants in the same context? The researcher attempts to account for changing conditions in the phenomenon chosen for study as well as changes in the design created by an increasingly refined understanding of the setting. The traditional quantitative view of reliability is based on the assumption that experiments can be replicated. Essentially, dependability is concerned with whether the same results would be obtained if observations could be made of the same thing twice. The idea of dependability emphasizes the need for the researcher to describe in detail the ever-changing context in which the research occurred and how these changes affected the way the research approached the study.

Confirmability

How can one be sure that the findings are reflective of the participants and inquiry themselves rather than a creation of the researcher's biases or prejudices? Confirmability captures the traditional concept of objectivity. Do the data help confirm the general findings and lead to the implications? Because qualitative research assumes that each researcher brings a unique perspective to the study, confirmability refers to the degree to which the results could be confirmed or corroborated by others. To enhance confirmability, the researcher can document the procedures for checking and rechecking the data throughout the study. Another researcher can review the results or the researcher can actively search for and describe negative instances that contradict prior observations. After the study, a data audit could be conducted that would examine the data collection and analysis procedures and make judgments about the potential for bias or distortion.

Many quantitative researchers see these criteria as a way to rename the quantitative criteria in order to increase the legitimacy of qualitative research. However, qualitative researchers believe that the traditional quantitative criteria cannot be applied to qualitative research. For example, how can the external validity of a qualitative study be judged if formalized sampling methods were not used? How can the reliability of qualitative data be judged when there is no mechanism for estimating the true score?

These alternative criteria remind researchers that qualitative research cannot be considered only an extension of the quantitative paradigm into the realm of nonnumeric data.

Analysis of Data

Content Analysis

Although I presented methods of data collection previously, this section will briefly describe ways that collected data can be analyzed. Content analysis is the analysis of text documents and can be quantitative, qualitative, or both. The purpose of content analysis is to identify patterns in text. Content analysis is an extremely broad area of research and includes thematic analysis of text, indexing, and quantitative descriptive analysis.

Thematic analysis of text is the identification of themes or major ideas in a document or set of documents. The documents can be any kind of text including field notes, newspaper articles, technical papers, or organizational memos. Researchers use computers to complete indexing. The computer program scans the text and indexes all key words. All key words are alphabetized and listed with the text that precedes and follows it so the researcher can see the word in the context in which it occurred in the text. The purpose of quantitative descriptive analysis is to describe features of the text quantitatively. A researcher might want to find out which words or phrases were used most frequently in the text.

When using content analysis, researchers are limited to the types of information available in text form. If the inquiry is about the way a news story is being handled by the news media, there would be a plethora of news stories from which to sample. On the other hand, if the research is on people's views on capital punishment, there may not be an appropriate archived text document. Researchers should also be careful when sampling in order to avoid bias. For example, a study on methods of treating AIDS might use the published literature as the population, thus excluding unpublished writing on AIDS and the most recent work that has not yet been published. Researchers also must be careful about interpreting results of automated content analyses because a computer program cannot infer meaning. It is easy in a large-scale analysis to misinterpret data because the subtleties of meaning were not taken into accounted. Content analysis has the advantage of being unobtrusive and can be an expedited method for analyzing large amounts of text.

Secondary Analysis of Data

Secondary analysis also uses archival data; however, secondary analysis typically refers to the reanalysis of quantitative data rather than text. The data that is routinely collected by governments, businesses, schools, and other organizations are stored in electronic databases that can be accessed and analyzed. Conveniently, many research projects store raw data in computer archives, making it easy for other researchers to analyze. Among the data available for secondary analysis are census bureau data, crime records, standardized testing data, economic data, and consumer data.

The greatest advantage of secondary analysis is that it is efficient. It makes use of data that were already collected by someone else, and it allows the researcher to extend the scope of the study. The costs involved in taking a national sample would keep many researchers from undertaking large studies. By using archived national databases, researchers can use a relatively small budget while producing a much broader study.

Difficulties with using secondary analysis are rooted in the data retrieval process. Accesses and linking data from large complex databases may require the assistance of computing professionals. By using data collected by others you are assuming they followed the correct procedures for data collection and storage. Large, well-financed national studies are usually documented quite thoroughly, but even detailed documentation of procedures is often no substitute for direct experience collecting data.

FUTURE DIRECTIONS

Because of the emphasis on in-depth knowledge and elaboration of images and concepts, qualitative methods have historically been useful for studying marginalized groups. Research in the fields of women, disability, education, social work, nursing, and human service has been and continues to be the focus of this type of inquiry. As qualitative researchers produce sophisticated studies using the methods described in this chapter, it will become more accepted to use qualitative methods to investigate more diverse individuals, groups, and social phenomena.

SUMMARY

Although using quantitative methods may seem easier, the data give you limited information on quantities and level of statistical significance (if any). Qualitative approaches, conversely, give you meanings and descriptions of personal significance. Which method a researcher decides to use depends upon the purposes and goals of research and is often some combination of both.

This chapter presented the basics of qualitative research to give beginning researchers basic knowledge and suggestions for finding more information about perspectives and methods that may be of further interest. There was an emphasis on the fact that there are different ways of understanding and knowing and that each researcher must identify the paradigm in which he or she operates prior to beginning research endeavors. It is from the chosen paradigm that researchers operate. What a researcher does will depend on the way the research questions were structured, the philosophical and methodological stance in which they function, the way the project was designed to support these, and the realities of the research process that has been pursued.

Successful researchers keep track of the pulse of the field. That is, they find out what other researchers with similar interests have done and are doing. Once a researcher has identified with a certain paradigm, he or she is encouraged to study other works and previous studies that have been done. Developing this foundation will give beginning researchers the basis upon which to build their own distinctive approach to qualitative inquiry.

REFERENCES AND FURTHER READINGS

Berg, B. L. (2004). *Qualitative research methods for the social sciences.* Boston: Pearson.

Berkwits, M., & Inui, T. (1998). Making use of qualitative research techniques. *Journal of General Internal Medicine, 13*, 195–199.

Blumer, H. (1969). *Symbolic interactionism: Perspective and methods.* Berkeley: University of California Press.

Bogdan, R. C., & Taylor, S. J. (1975). *Introduction to qualitative methods: A phenomenological approach to the social sciences.* New York: Wiley.

Bremme, D. W., & Erikson, F. (1977). Relationships between verbal and non-verbal classroom behaviors. *Theory and Practice, 16,* 153–161.

Cole, J. (1983). The value of a person lies in his Herzensbildung: Franz Boas' Baffin Island letter-diary, 1883–1884. In G. Stocking (Ed.), *Observers observed* (pp. 13–52). Madison: University of Wisconsin Press.

Denzin, N. K. (1989). *Interpretive biography.* Newberry Park, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (1994). Introduction: Entering the field of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1–18). Thousand Oaks, CA: Sage.

Erikson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York: Macmillan.

Gilligan, C. (1983). *In a different voice: Psychological and women's development.* Cambridge, MA: Harvard University Press.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine De Gruyther.

Goodall, H. L. (2000). *Writing the new ethnography.* Lanham, MD: AltaMira.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 60–69). Thousand Oaks, CA: Sage.

Hatch, J. A. (2002). *Doing qualitative research in educational settings*. Albany: State University of New York Press.

Inui, T., & Frankel, R. M. (1991). Evaluating the quality of qualitative research: A proposal pro tem. *Journal of General Internal Medicine, 6,* 485–486.

Jackson, P. (1968). *Life in classrooms.* New York: Holt, Rinehart, & Winston.

Krueger, R. A. (1988). *Focus groups: A practical guide for applied research.* Newberry Park, CA: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Mason, J. (2002). *Qualitative researching.* London: Sage Publications.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* (2nd ed.). Newbury Park, CA: Sage.

Rosaldo, R. (1989). *Culture and truth: The remaking of social analysis.* Boston: Beacon.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newberry Park, CA: Sage.

Vidich, A. J., & Lyman, S. M. (1994). Qualitative methods: Their history in sociology and anthropology. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1–18). Thousand Oaks, CA: Sage.

12

Ethics of Psychological Research

Elizabeth V. Swenson
John Carroll University

To do the best research and to give the best service to the community and to the profession, investigators need to behave ethically. Much has been written about the ethics of research in the behavioral sciences. For research in psychology specifically, the guidelines are set forth clearly in Section 8, Research and Publication, of the American Psychological Association (APA) Ethical Principles of Psychologists and Code of Conduct (Ethics Code; APA, 2002). It was specifically designed to meet "ethical challenges in the new millennium" (Fisher, 2003, p. xxv). Childress, Meslin, and Shapiro (2005) call the APA Ethics Code the best-known ethics code for the conduct of behavioral and social research (p. 265). Although it can be supplemented by other documents, the Ethics Code is the foundation for this chapter on the ethics of psychological research.

The Ethics Code consists of 10 sections. Within Section 8 there are 15 standards, the first 9 of which specifically refer to research, with the last 6 covering publication of research results. Of these, only one refers to psychological research with animal subjects. The standards are preceded by five general ethical principles that underlie the standards in the code and are aspirational in nature: beneficence and nonmaleficence, fidelity and responsibility, integrity, justice, and respect for people's rights and dignity (see Figure 12.1).

These principles are considered to be the moral basis for the Ethics Code, and are similar to those well known in bioethics (Beauchamp & Childress, 2001).

In applying the standards of the Ethics Code to specific situations in research, often the correct answer is not readily apparent. The investigator then needs to apply the general ethical principles to aid in decision making, after which she should consult a colleague and then document the process.

HISTORY

The protection of human participants[1] is a huge issue in research. It was not always so.

One need only look to the egregious use of Nazi concentration camp prisoners for biomedical "experiments" during World War II or to the poor African American men with untreated syphilis recruited by the United States government for a study of the illness in the 1940s to understand that the rights of research participants have been frequently and flagrantly ignored. The Nuremberg Code, written during the Nuremberg War Crimes trial, is considered to be the first code of ethics for research

[1] The fourth edition of the *Publication Manual of the American Psychological Association* (APA, 1994) was the first to recommend that human subjects in an experiment be called "participants" instead, as a more humanizing referent. This practice is followed in this chapter.

Table 12.1 The general principles from the Ethics Code that guide and inspire psychologists in all their work

Principle A: Beneficence and Nonmaleficence

Psychologists strive to benefit those with whom they work and take care to do no harm. In their professional actions, psychologists seek to safeguard the welfare and rights of those with whom they interact professionally and other affected persons, and the welfare of animal subjects of research.

Principle B: Fidelity and Responsibility

Psychologists establish relationships of trust with those with whom the work. They are aware of their professional and scientific responsibilities to society and to the specific communities in which they work. Psychologists uphold professional standards of conduct.

Principle C: Integrity

Psychologists seek to promote accuracy, honesty, and truthfulness in the science, teaching, and practice of psychology. In these activities psychologists do not steal, cheat, or engage in fraud, subterfuge, or intentional misrepresentation of fact….In situations in which deception may be ethically justifiable to maximize the benefits and minimize harm, psychologists have a serious obligation to consider the need for, the possible consequences of, and their responsibility to correct any resulting mistrust or harmful effects that arise from the use of such techniques.

Principle D: Justice

Psychologists recognize that fairness and justice entitle all persons to access to and benefit from the contributions of psychology.…Psychologists exercise reasonable judgment and take precautions to ensure that their potential biases…and the limitations of their expertise do not lead or condone unjust practices.

Principle E: Respect for People's Rights and Dignity

Psychologists respect the dignity and worth of all people, and the rights of individuals to privacy, confidentiality, and self-determination. Psychologists are aware that special safeguards may be necessary to protect the rights and welfare of persons or communities whose vulnerabilities impair autonomous decision making. Psychologists are aware of and respect cultural, individual, and role differences, including those based on age, gender, gender identity, race, ethnicity, culture, national origin, religion, sexual orientation, disability, language, and socioeconomic status.

SOURCE: APA, 2002.

with human participants. See Capron (1989) to review the Nuremberg Code.

In 1974 Congress passed the National Research Act, which created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. It consisted of 11 people who were charged with developing ethical principles and guidelines for research involving human participants. This commission produced the Belmont Report (1978) in which the principles of justice, beneficence, and respect for persons (autonomy) were set forth. Sherwin (2005) notes that the Belmont Report "cites the 'flagrant injustice' of the Nazi exploitation of unwilling prisoners and the Tuskegee syphilis study as evidence of the shameful history of research abuses" (p. 151). According to Jonsen (2005), "If research involving human persons as subjects is to be appraised an as ethical activity it must above all be an activity in to which persons freely enter" (p. 7). Thus the central importance of informed consent was established.

The first APA Ethics Code was published in 1953. Throughout its 10 revisions, the goal has been to be to define the moral standards and values that unite us as a profession and a discipline that both treats and studies behavior.

THE INSTITUTIONAL REVIEW BOARD

To protect the human participants in research, an institution that applies for federal research funds is required to set up an Institutional Review Board (IRB; United States Department of Health and Human Services [DHHS], 1991). The IRB must consist of at least five researchers/faculty from a variety of disciplines who can review research at the institution with at least one member of the community who represents the public. Its mission is to ensure that human participants are dealt with in an ethical manner.

> The Ethics Code references the IRB in Standard 8.01.
> 8.01 Institutional Approval
> When institutional approval is required, psychologists provide accurate information about their research proposals and obtain approval prior to conducting the research. They conduct the research in accordance with the approved research protocol.

Although some minimally risky pro forma class activities can be reviewed at the class or department level, any research project with risks to participants or which deals with vulnerable individuals such as minors or incarcerated people must have a full IRB review before commencing the research. Minimal risk is usually thought of as the risk inherent in daily living. The IRB reviews the investigator's proposal with special attention to the experimental procedure, the informed consent process, the description and recruitment of the participants to be used, and the rationale for the research. The IRB then approves or disapproves of the research or provides information that can be used to modify the procedures proposed to more fully protect the participants.

For example, Martha's senior thesis project is investigating how hearing a background of happy or sad music might affect person's prediction of how well a pictured individual will do in a job interview. It appears that these activities are minimally risky; people listen to music and make implicit predictions every day. The college's IRB reviewed this proposal. One member wondered what might happen if a participant had had a recent breakup with a boyfriend and was reminded of this event by the sad music. The IRB then recommended a routine statement in the debriefing that if one is affected adversely by the experiment one might visit the university counseling center. This comment was accompanied by the telephone number and e-mail of the center along with the routine contact information of the principle investigator, the investigator's faculty sponsor, and the IRB.

INFORMED CONSENT

Informed consent must be given knowingly and voluntarily by a competent person to be able to participate in psychological research. The three requirements are therefore information, understanding, and voluntariness. The first of these three requirements is apparent from reading Standard 8.02.

> 8.02 Informed Consent to Research
> (a) When obtaining informed consent as required in Standard 3.10, Informed Consent, psychologists inform participants about (1) the purposes of the research, expected duration, and procedures; (2) their right to decline to participate and to withdraw from the research once participation has begun; (3) the foreseeable consequences of declining or withdrawing; (4) reasonably foreseeable factors that may be expected to influence their willingness to participate such as potential risks, discomfort, or adverse effects; (5) any prospective research benefits; (6) limits of confidentiality; (7) incentives for participation; and (8) whom to contact for questions about the research and research participants' rights. They provide opportunity for the prospective participants to and questions and receive answers.

Standard 8.02(a) list sets forth all of the information that a prospective participant is entitled to know about the research before making a decision about whether to participate. She should also have a clear understanding about what she will be asked to do in the study. This is particularly important if the research participants are to be students who either have a relationship with the investigator or are participating to obtain research points for a class assignment. Often introductory psychology students have an assignment to participate in a number of psychology experiments. For these students, having a definite understanding of the right to withdraw with no questions asked or to decline to participate at all, along with the right to collect any compensation, is crucial to the voluntary nature of informed consent. Since many students are skeptical about the procedures psychologists use to conduct research, a full explanation of risks involved, including mental stress is most important. Time to ask questions is needed to ensure full understanding.

The limits to confidentiality are also an important factor to be discussed during the informed consent period. Here Standard 8.02 is supplemented by Standard 4.01.

> Standard 4.01 Maintaining Confidentiality
> Psychologists have a primary obligation and take reasonable precautions to protect confidential information obtained through or stored in any medium, recognizing that the extent and limits of confidentiality may be regulated by law or established by institutional rules or professional or scientific relationship.

Participants may be asked to answer questions about their childhood, mental health, sexual orientation, or illegal activities, for example. This is obvious information to be kept confidential. If questions arise about the possibility of suicide and they are answered in a manner predictive of risk, an investigator might be obligated not to keep this information confidential. Contingency plans for this possible occurrence should be made with the college counseling center, for example. Indications that a child is likely being abused likewise need prompt attention. If a participant admits that she was abused and that she has younger siblings still at home, she needs to be aware beforehand that this is beyond the limits of confidentiality and must be reported to the local children and family services agency. It is recommended that participants be identified by using a random number so that they do not wonder if their data are recognizable or traceable.

Often anonymity is assumed because participants do not sign their name on a questionnaire. But any identifying questions or demographic information requested might make it possible to identify a particular person. For example, a research project is proposed where participants, students in the introductory psychology course, will be anonymous. Participants are asked to fill out a form indicating their age, sex, and race. Few students of nontraditional age or minority status are in the course, making it relatively easy to identify the person who says his age is 42.

Concerning the voluntary aspect of informed consent, it is important for participants not to feel any pressure to agree. Understanding the foreseeable consequences of withdrawing lessens this pressure. In addition, fully understanding the incentives for participation leads to a more rational decision.

Finally, participants need to be competent to consent to research. Competence is closely related to understanding, but adds a different twist. In the legal system, a person under the age of 18 is considered to be a juvenile. As such, a 17-year-old would not be considered competent to consent to being a research participant in most instances. To err on the cautious side, it would be unwise to ask a college student or anyone else under 18 to give informed consent, even though the person, if a college student, may be able to do so.

Table 12.2 A sample informed consent form

Name___________________________________

Research project title_______________________

Principal investigator_________________

I understand and have had an opportunity to ask questions about: (initial each statement)

the purpose of the research, expected duration, and procedures.______

my right to decline to participate and to withdraw from the research at any point once participation has begun.______

any foreseeable consequences of my declining or withdrawing.______

any potential risks, discomfort, or adverse effects.______

any prospective research benefits to me for participating.______

any limits of confidentiality.______

any incentives for participating.______

any use of audio or video recording.______

whom to contact for questions about the research and my rights as a participant.______

I freely agree to participate in this experiment.______

I am over 18 years old.______

___________________________(signed) ____________(date)

Informed consent should always be documented. This is best done by asking the participants to sign a statement of understanding, although an electronic signature is acceptable. Alternatively, the investigator could document that the participant understood the statement and verbally agreed to participate.

Table 12.2 illustrates a sample informed consent form that covers all the bases.

A general informed consent standard from the Ethics Code, 3.10, states some additional considerations when seeking informed consent.

> Standard 3.10 Informed Consent
>
> (a) When psychologists conduct research or provide assessment, therapy, counseling, or consulting services in person or via electronic transmission or other forms of communication they obtain informed consent of the individual or individuals using language that is reasonably understandable to that person or persons except when conducting such activities without consent is mandated by law or governmental regulation or as otherwise provided in this Ethics Code.
>
> (b) For persons who are legally incapable of giving informed consent, psychologists nevertheless (1) provide an appropriate explanation, (2) seek the individual's assent, (3) consider such persons' preferences and best interests, and (4) obtain appropriate permission from a legally authorized person, if such substitute consent is permitted or required by law. When consent by a legally authorized person is not permitted or required by law, psychologists take reasonable steps to protect the individual's rights and welfare.
>
> (c) When psychological services are court ordered or otherwise mandated, psychologists inform the individual of the nature of the anticipated services, including whether the services are court ordered or mandated and any limits of confidentiality, before proceeding.
>
> (d) Psychologists appropriately document written or oral consent, permission, and assent.

In Standard 3.10 (a), the investigator is instructed to ask for informed consent "using language that is reasonably understandable" to the person. This would entail simplifying language that is more sophisticated or full of jargon than a person can be seen to understand. In addition, it would require some one who has less facility with English to be presented with the agreement in a native language, if appropriate.

The code also speaks to additional requirements for those potential participants who are legally incapable of consenting or legally incompetent to consent. If we consider this potential participant to be a 15-year-old adolescent, the code instructs the investigator to explain the study to the person, using reasonably understandable language. A person of this age could probably process all of the elements required by the code in Standard 8.02(a), so these would need discussing with the adolescent. Then the investigator would ask her if she would be willing to assent (agree) to being a participant, after first determining if it would be in her best interest to participate. Finally, the formal informed consent procedure outlined in Standard 8.02(a) would be presented to the parent or guardian. This same procedure would apply if the participant was a legally incompetent adult.

For example, John is interested in studying bullying behavior in middle school boys. He develops three kinds of questionnaires to be presented to the teacher, the child, and the parent. He prepares informed consent forms for each person which use the same words to describe the study and only differ in the words "my child cannot participate" for the parents, "I will participate in rating my students" for the teacher, and "I would like to participate" for the children. For the parents, he noted that they need sign the permission only if they did not want their child to participate. The IRB spent some time deciphering exactly what John was planning to do and whether this research plan could be modified to meet ethical standards. They questioned whether they should go so far as to redesign the study for John, or just disapprove the study. They decided that for this time they would only deal with the informed consent problems. They told John that the parents must sign an affirmative permission, not give it by default. They recommended that the permission slips not be sent home with the students or returned by the students for the possibility that this would put pressure on the students. They also disapproved of the wording of the permissions, feeling that many of the parents may not be able to read at the 12th grade level the permissions were written in. In addition, John did not appear to understand that the assent forms should be discussed with each student individually to insure understanding. They asked John to think about how a study of bullying behavior conducted by a college student might have some adverse impact on the child participants.

Distinctive Instances of Informed Consent

Intervention Research

When research is conducted in new kinds of treatment for mental health problems, special considerations apply. These are set forth in Standard 8.02(b).

> 8.02(b) Psychologists conducting intervention research involving the use of experimental treatments clarify to participants at the outset of the research (1) the experimental nature of the treatment; (2) the services that will or will not be available to the control group(s) if appropriate; (3) the means by which assignment to treatment and control groups will be made; (4) available treatment alternatives if an individual does not wish to participate in the research or wishes to withdraw once a study has begun; and (5) compensation for or monetary costs of participating including, if appropriate, whether reimbursement from the participant or a third-party payor will be sought.

Thus, participants need to know that the treatment is experimental in nature, how the control group will be treated or not treated, how each individual is assigned to the treatment or to the control group, alternatives to the experimental treatment if the individual does not want to participate, and compensation issues. These additional components of the informed consent are particularly important because there is a likely possibility that the participants are from a vulnerable population, could easily be taken advantage of or exploited by the therapist/investigator, and might mistakenly believe that an experimental treatment is an improvement over the usual treatment.

Recording

Occasionally it is necessary to record (either audio or video) the behavior of a research participant. A participant has the right to know of the recording in advance and to decline participation. Standard 8.03 makes two exceptions.

> 8.03 Informed Consent for
> Recording Voices and Images in Research
> Psychologists obtain informed consent from research participants prior to recording their voices or images for data collection unless (a) the research consists solely of naturalistic observations in public places, and it is not anticipated that the recording will be used in a manner that could cause personal identification or harm, or (b) the research design includes deception, and consent for the use of the recording is obtained during debriefing.

Special Research Participants

> 8.04 Client/Patient, Student,
> and Subordinate Research Participants
> (a) When psychologists conduct research with clients/patients, students, or subordinates as participants, psychologists take steps to protect the prospective participants from adverse consequences of declining or withdrawing from participation.
> (b) When research participation is a course requirement or an opportunity for extra credit, the prospective participant is given the choice of equitable alternative activities.

Special care needs to be given to students, clients, or subordinates who participate in research. As noted above, precautions need to be taken to be sure that no pressure is placed on these individuals to participate. Standard 8.04(a) states that steps must be taken to protect students, for example, from a decision to withdraw or decline.

Standard 8.04(b) adds a further requirement. The question here is what is an equitable alternative activity to the course requirement? College students in the introductory psychology class are frequently used as research participants because they are convenient and plentiful. Their use is normally justified, however, by the educational value of learning about psychological research firsthand. An equitable alternative activity should take about the same amount of time and also result in learning about research. Possible examples include reading about how to design and carry out experiments, reviewing articles on empirical research, or even helping out with data collection. Most likely the students who actually do these alternative activities are those who run out of time to be participants or who need to do assignments on their own schedule.

For example, in Maggie Smith's Psychology 101 class, students are required to participate in four 30-minute units of research. Alternatively, they may elect to review four research articles from the psychology literature. For this assignment students must locate an article and describe the participants, the procedure, and the results. Maggie decides that even though she dislikes the idea of participating in research because it disrupts her carefully planned child care schedule, she must participate because the reviewing of four articles is more difficult and much more time consuming. She does not give her informed consent voluntarily.

When Informed Consent Is Not Required

Standard 8.05 of the Ethics Code states,

> 8.05 Dispensing With Informed Consent for Research
> Psychologists may dispense with informed consent only (1) where research would not reasonably be assumed to

> create distress or harm and involves (a) the study of normal educational practices, curricula, or classroom management methods conducted in educational settings; (b) only anonymous questionnaires, naturalistic observations, or archival research for which disclosure of responses would not place participants at risk of criminal or civil liability or damage their financial standing, employability, or reputation, and confidentiality is protected; or (c) the study of factors related to job or organization effectiveness conducted in organizational settings for which there is no risk to participants' employability, and confidentiality is protected or (2) where otherwise permitted by law or federal or institutional regulations.

For example, Sally wanted too know if preschool boys played more roughly than preschool girls. She went to a park playground where day care teachers took their children for their daily recess play. She unobtrusively tallied the number of rough play incidents between boys and boys, between girls and girls, and between boys and girls. For these frequency data taken in a naturalistic observation no informed consent was needed.

Ignoring informed consent in school settings should be done very cautiously. Interventions or assessments that are not a routine part of the education of children should be consented to by a parent or guardian, although at times a school administrator will assume this responsibility on behalf of the children and parents.

Inducements

The voluntariness of consent can be affected by the inducements offered to participate. Typical inducements can be money, gift certificates, research points for a class requirement, extra-credit points for a class, and professional service. Excessive inducements can pressure a person to consent and affect its voluntariness.

> 8.06 Offering Inducements for Research Participation
> (a) Psychologists make reasonable efforts to avoid offering excessive or inappropriate financial or other inducements for research participation when such inducements are likely to coerce participation.
> (b) When offering professional services as an inducement for research participation psychologists clarify the nature of the services, as well as the risks, obligations, and limitations.

For example, if participants are recruited from students living in two college dorms, would $10 for an hour of participation be excessive? What about $20?

DEBRIEFING

Standard 8.08 states,

> 8.08 Debriefing
> (a) Psychologists provide a prompt opportunity for participants to obtain appropriate information about the nature, results, and conclusions of the research, and they take reasonable steps to correct any misconceptions that participants may have of which the psychologists are aware.
> (b) If scientific or humane values justify delaying or withholding this information, psychologists take reasonable measures to reduce the risk of harm.
> (c) When psychologists become aware that research procedures have harmed a participant, they take reasonable steps to minimize the harm.

Researchers ideally talk with participants immediately after the conclusion of their participation, giving information about the experiment and answering questions. This is what makes research participation truly educational. In addition, investigators can learn from participants about difficulties they may have had in understanding directions, for example. Research results can rarely be given at this time so researchers notify participants of where they can learn of the results and conclusions. This might mean sending participants an abstract of the results, or at least offering to do so, or posting a description of the study along with the results and conclusions on a bulletin board or Web site that participants can access.

Often a particular participant may be upset with an experiment. Procedures that would be quite neutral for some may prove to be disturbing for others. Should this be the case, the Code requires that the investigator take "reasonable" care to alleviate the harm. What is reasonable varies from one person to another and is difficult to define. In fact, it has a legal meaning that a jury of ones peers would find an act to be reasonable. Because it is unlikely, hopefully so, that a jury would ever be involved, a good substitute is to consult a colleague to confirm that one's actions appear reasonable. This consultation should always be documented along with the informed consent. For most idiosyncratic reactions to an experimental procedure, referral to the university counseling center or a mental health clinic just to talk with someone is often recommended.

For example, Steve was asked to visualize and then view a movie of a crime committed by a mentally disturbed person. During the debriefing, it became apparent that the movie was deeply disturbing to Steve. The investigator, who had foreseen this possibility, discussed with Steve the arrangements that had been made for participants to debrief further with a psychology intern in the department.

THE SPECIAL CASE OF DECEPTION IN RESEARCH

Can a research participant ever give informed consent to be in a study where deception is involved? Reflect on this while reading Ethics Code Standard 8.07.

> 8.07 Deception in Research
> (a) Psychologists do not conduct a study involving deception unless they have determined that the use of deceptive techniques is justified by the study's significant prospective

> scientific, educational, or applied value and that effective nondeceptive alternative procedures are not feasible.
> (b) Psychologists do not deceive prospective participants about research that is reasonably expected to cause physical pain or severe emotional distress.
> (c) Psychologists explain any deception that is an integral feature of the design and conduct of an experiment to participants as early as is feasible, preferably at the conclusion of their participation, but no later than at the conclusion of the data collection and permit participants to withdraw their data.

Two types of deception study are the most common in psychological research. One type of deception study involves a confederate of the experimenter whom the participant believes to be someone else, perhaps a fellow participant. In the second instance, the participant is given false or misleading feedback about his/her work.

Consequences to the participant can range widely. Many participants, upon being dehoaxed or debriefed, have no reaction or are amused by the disclosure. Others feel duped, deceived, like they've been made a fool of, or emotionally harmed by the experience. Some feel that they cannot trust a psychologist/experimenter again. Some experience a loss of self-esteem. If some participate in another experiment, they try to figure out the real purpose of the research and thus are not naïve participants again. The reputation of psychological research can be damaged in the eyes of participants.

Some deception research is relatively harmless whereas other studies can leave lasting effects on the participants. Consider the following two deception studies, the first of which is a classic involving a confederate of the experimenter, and the second, a more recent study, giving false feedback to the participants.

Some citizens of New Haven, Connecticut elected to participate in the obedience study conducted by Stanley Milgrim (1963). These participants were told that they were part of a study on the effects of electric shock on learning. They were not told that the real purpose of the study was to see how far a person would go in shocking a difficult learner. The "learner," who just couldn't learn, was actually a confederate of the experimenter, acting increasingly harmed by the "electric shock." Surprisingly, the researchers concluded that a majority of everyday people would shock another person to the point of what they thought to be great injury, all in the name of science or in obedience to a researcher in a white lab coat.

Compare the Milgrim study with one over 30 years later. Kassin and Kiechel (1995) instructed participants to type letters read to them as fast as they could. They were instructed not to touch the "alt" key as this would spoil the experiment. In one condition participants were told that they had in fact touched the forbidden key and ruined the experiment. Several of these participants not only felt that they had destroyed the data but made up an explanation for how they had done so. The researchers concluded that, given the right conditions of interrogation, people can fabricate confessions and even come to believe them.

Applying Standard 8.07(a), it is first necessary to look to the significance of the scientific, educational, or applied value of the study. Assuming that the studies are of significance, it is required next to look at whether effective nondeceptive alternative procedures are feasible. Although this is controversial, assume that the two studies are of significance and that no nondeceptive procedures could have yielded these results. In Standard 8.07(b) we are instructed to contemplate whether participation in these projects could reasonably be expected to cause physical pain or emotional distress. Most people would agree that deciding to shock another person to the point of great physical injury is many steps above accidentally ruining an experiment. Today, it is unlikely that Milgrim could have had his study approved by his university's IRB.

Finally, Standard 8.07(c) provides a remedy for the participant in a deception study. Upon dehoaxing, the participant may request that her data be withdrawn from the study. Thus she is saying that if I had been informed of the true procedure of the study, I would not have consented to being a participant.

Does this amount to informed consent? Note that the researcher in the debriefing does not need to offer to let the participant withdraw her data. The participant needs to come up with this idea on her own. In contrast, the Canadian Code of Ethics for Psychologists (Canadian Psychological Association, 2000, pp. 23, 26) states

> In adhering to the Principle of Integrity in Relationships, psychologists would…
>
> III.29 Give a research participant the option of removing his or her data, if the research participant expresses concern during the debriefing about the incomplete disclosure or the temporary leading of the research participant to believe that the research project or some aspect of it had a different purpose, and if the removal of the data will not compromise the validity of the research design and hence diminish the ethical value of the participation of the other research participants.

In the Canadian code, the researcher is to give the option of withdrawing data to the participant. With respect to the validity of the experiment, the argument can be made that this participant would not have consented to being in the experiment if she had known its true nature.

Another possible solution to the informed consent to deception research may be feasible in a setting where there is a very large subject pool. Prior to selection for any particular research participation, individuals put themselves in to one of two categories: those who are willing to participate in deception research and those who are not. Presumably those in the first category are on their guard for being deceived, making them somewhat more suspicious of the true nature of the research than those in the second category.

In addition, Veatch (2005) notes in reference to Standard 8.07 that "the American Psychological Association has long endorsed behavioral research on terms that violate both the Department of Health and Human Services regulations and the spirit of Belmont" (p. 189). Specifically this refers to the principle of autonomy in the Belmont Report and the precautions found in the DHHS (1991) regulations. Autonomy, the right to make ones own decision about participating in research, is arguably not respected in deception research.

RESEARCH WITH ANIMALS

Much information that is useful for the understanding of human behavior or for the welfare of humans comes from research with animals. The utility of the research, however, does not justify the unethical treatment of the animal subjects. Many feel that because animals cannot ever give informed consent to research and may be physically injured by the research, very special care should be used to treat animals ethically if they must be used in research at all. Analogous to the IRB, the federal government also requires institutions using animals in research to form an Institutional Animal Care and Use Committee (IACUC) to oversee the ethical treatment of research animals, including their acquisition and housing.

The APA Ethics Code states,

> 8.09 Humane Care and Use of Animals in Research
> (a) Psychologists acquire, care for, use, and dispose of animals in compliance with current federal, state, and local laws and regulations, and with professional standards.
> (b) Psychologists trained in research methods and experienced in the care of laboratory animals supervise all procedures involving animals and are responsible for ensuring appropriate consideration of their comfort, health, and humane treatment.
> (c) Psychologists ensure that all individuals under their supervision who are using animals have received instruction in research methods and in the care, maintenance, and handling of the species being used, to the extent appropriate to their role.
> (d) Psychologists make reasonable efforts to minimize the discomfort, infection, illness, and pain of animal subjects.
> (e) Psychologists use a procedure subjecting animals to pain, stress, or privation only when an alternative procedure is unavailable and the goal is justified by its prospective scientific, educational, or applied value.
> (f) Psychologists perform surgical procedures under appropriate anesthesia and follow techniques to avoid infection and minimize pain during and after surgery.
> (g) When it is appropriate that an animal's life be terminated, psychologists proceed rapidly, with an effort to minimize pain and in accordance with accepted procedures.

Standard 8.09 is supplemented by the APA Guidelines for Ethical Conduct in the Care and Use of Animals (APA, 1997)

For example, Harriett, a university work/study student has been employed by Dr. Green in the neuroscience department to care for his laboratory rats over the weekends. She has been carefully trained by Dr. Green on the procedures to be used. At the last minute, Harriett has been given the opportunity to spend a weekend in New York City with a friend. She searches around for someone who will take care of the animals while she is gone. Robert, one of her classmates, agrees to feed the animals. Without knowledge of the feeding procedures, Robert feels that he can feed the animals double on Saturday and omit the Sunday feeding. While measuring out the food, Robert finds that the temperature in the lab is too warm for comfort, so he turns the thermostat down as far as it will go, fully intending to turn it back up when he leaves. On Monday, Dr. Green tells Harriett that she has spoiled his research with her carelessness.

ETHICS CODE STANDARDS THAT DEAL WITH REPORTING RESEARCH RESULTS

Standard 8.10 makes a definitive statement about the ethics of the dissemination of research results.

> 8.10 Reporting Research Results
> (a) Psychologists do not fabricate data.
> (b) If psychologists discover significant errors in their published date, they take reasonable steps to correct such errors in a correction, retraction, erratum, or other appropriate publication means.

In very few standards of the Ethics Code is there a statement that absolutely prohibits an action. The fabrication of data is one of those cases. If an error is make in the publication of data, psychologists have an obligation to publish a retraction in whatever form is appropriate for the purpose. The Ethics Code makes definite the unethicality of any kind of scientific misconduct.

In Standard 8.11, there is another absolute prohibition.

> 8.11 Plagiarism
> Psychologists do not present portions of another's work or data as their own even if the other work or data source is cited occasionally.
> Plagiarism is a very serious offense. The taking of another's words or ideas as ones own is the theft of intellectual property. This unethical usage also extends to the oral use of another's words or ideas without crediting the originator. This is known as oral plagiarism.
> 8.12 Publication Credit
> (a) Psychologists take responsibility and credit, including authorship credit, only for work they have actually performed or to which they have substantially contributed.
> (b) Principle authorship and other publication credits accurately reflect the relative scientific or professional contributions of the individuals involved, regardless of their relative status. Mere possession of an institutional position, such as department chair, does not justify authorship credit. Minor contributions to the research or to the writing for publications are acknowledged appropriately, such as in footnotes or in an introductory statement.

(c) Except under exceptional circumstances, a student is listed as principal author on any multiple-authored article that is substantially based on the student's doctoral dissertation. Faculty advisors discuss publication credit with students as early as feasible and throughout the research and publication process as appropriate.

Although the order of names on a multiple-authored article may seem unimportant, it is just the opposite. Those who are in a position of evaluating a person's work, assume that the first stated author did more of the scholarly work than the second author did, and so forth. Normally, if each author's contribution was similar, this fact is mentioned in a footnote to the authorship. It is also assumed that because a dissertation represents an original work of scholarship, an article based on a person's dissertation will list that person's name first or solely. Contributions to a publication in the form of collecting dates, running a computer program, or assisting with a literature review are normally given a footnote credit.

8.13 Duplicate Publication of Data
Psychologists do not publish, as original data, data that have been previously published. This does not preclude republishing data when they are accompanied by proper acknowledgment.
8.14 Sharing Research Data for Verification
(a) After research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professions who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude psychologists from requiring that such individuals or groups be responsible for costs associated with the provision of such information.
(b) Psychologists who request data from other psychologists to verify the substantive claims through reanalysis may use shared data only for the declared purpose. Requesting psychologists obtain prior written agreement for all other uses of the data.

Together Standards 8.13 and 8.14 deal with other aspects of scientific honesty. Data that are published more than once without a proper acknowledgment erroneously suggest that multiple data collections and analyses have been done. Sharing ones published data with a colleague for reanalysis is not only a professional courtesy, but also helps to ensure that the data have been legitimately collected and carefully and accurately analyzed. Clearly, one who requests data to reanalyze must not use these data as her own for research and publication purposes. A similar standard, 8.15, applies to reviewers for publication.

8.15 Reviewers
Psychologists who review material submitted for presentation, publication, grant, or research proposal review respect the confidentiality of and the proprietary rights in such information of those who submitted it.

THE FUTURE OF RESEARCH ETHICS IN THE 21ST CENTURY

The general ethical principles that underlie the Ethics Code also provide foundational stability to the values of the profession and the discipline of psychology. It is unlikely that the 21st century will produce any more important values for psychologists than doing good, not harm; being trustworthy; promoting integrity and justice and respecting the rights and dignity of all people.

Ten versions of the Ethics Code were adopted between 1953 and 2002. With this background, one could confidently predict that the Ethics Code will be revised approximately nine times in the 21st century. Likely events to bring about such a modification would be changes in federal law, advances in technology, and political and social issues all of which can raise unforeseen problems with standards in the Ethics Code.

REFERENCES AND FURTHER READINGS

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

American Psychological Association. (1996). *Guidelines for ethical conduct in the care and use of animals.* Washington, DC: Author. Retrieved January 1, 2007, from http://www.apa.org/science/anguide.htm

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073.

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics* (5th ed.). New York: Oxford University Press.

Canadian Psychological Association. (2000). *Canadian Code of Ethics for Psychologists* (3rd ed.). Ottawa, Ontario, Canada: Author.

Capron, A. M. (1989). Human experimentation. In R. M. Veatch (Ed.), *Medical ethics* (pp. 125–172). Boston: Jones and Bartlett.

Childress, J. F., Meslin, E. M., & Shapiro, H. T. (Eds.). (2005). *Belmont revisited: Ethical principles for research with human subjects.* Washington, DC: Georgetown University Press.

Fisher, C. B. (2003). *Decoding the ethics code: A practical guide for psychologists*. Thousand Oaks, CA: Sage.

Jonsen, A. R. (2005). On the origins and future of the Belmont Report. In J. F. Childress, E. M. Meslin, & H. T. Shapiro (Eds.), *Belmont revisited: Ethical principles for research with human subjects*. Washington, DC: Georgetown University Press.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science, 7,* 125–128.

Korn, J. H. (1997). *Illusions of reality: A history of deception in social psychology*. Albany: State University of New York Press.

LeBlanc, P. (2001, September). "And mice…" (or tips for dealing with the animal subjects review board. *APS Observer, 14*, 21–22.

McCallum, D. M. (2001, May/June). “Of men…” (or how to obtain approval from the human subjects review board). *APS Observer, 14*, 18–29, 35.

Milgrim, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67,* 371–378.

Miller, N. E. (1985). The value of behavioral research on animals. *American Psychologist, 40*, 423–440.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research* (DHEW Publication OS No. 78-0012). Washington, DC: U.S. Government Printing Office.

Sherwin, S. (2005). Belmont revisited through a feminist lens. In J. F. Childress, E. M. Meslin, & H. T. Shapiro (Eds.), *Belmont revisited: Ethical principles for research with human subjects*. Washington, DC: Georgetown University Press.

United States Department of Health and Human Services. (1991). Federal policy for the protection of human subjects: Notices and rules. *Federal Register, 46*(117), 28001–28032

Veatch, R. M. (2005). Ranking, balancing, or simultaneity: Resolving conflicts among the Belmont principles. In J. F. Childress, E. M. Meslin, & H. T. Shapiro (Eds.), *Belmont revisited: Ethical principles for research with human subjects.* Washington, DC: Georgetown University Press.

PART III

NEUROSCIENCE

13

Biological Psychology

Lewis Barker
Auburn University

Ron, a married, 40-year-old schoolteacher, had a good life until it went bad. Nothing seemed to account for the overwhelming sexual urgency he began to experience—from searching Web sites for child pornography to soliciting prostitutes at massage parlors. Even though he knew that what he was doing was wrong, he lamely explained that the "pleasure principle overrode [his] restraint." Ron's wife found out about his aberrant behavior, had him evicted, and then prosecuted for pedophilia (fondling children). He was medicated and remanded to treatment for sex offenders. But he put himself at risk for a jail sentence when he violated the terms of treatment by soliciting sex from other program participants (Choi, 2002).

The night before sentencing, Ron went to an emergency room complaining of headaches and uncontrollable sexual urges. The attending physician referred him to the neurology division. There neurologists determined that he had a problem maintaining his balance and ordered an MRI scan. Further behavioral tests found that he couldn't write or copy drawings and that he didn't become particularly upset when he urinated on himself!

The MRI revealed an egg-sized brain tumor in the right side of Ron's forebrain—specifically, in his orbitofrontal cortex. From the study of other patients with a damaged orbitofrontal cortex, the neurologists knew that poor impulse control, errors of judgment, and problems with social behavior were common outcomes. And because most people diagnosed with pedophilia experience their problems throughout a lifetime, yet Ron's problems were recent ones, the neurologists confidently connected his behavioral disorder to the tumor.

Surgeons removed the tumor. However, seven months later, Ron's headaches returned, and again he secretly began to collect pornography. Perhaps his problem was, as most of society would think, a moral failing, unrelated to how his brain functioned. But another MRI revealed another tumor, and another surgery to remove it successfully alleviated the symptoms (Choi, 2002). To the extent Ron had a moral problem, could it be attributed to a misplaced tumor?

A BIOLOGICAL UNDERSTANDING OF MIND AND BEHAVIOR

Like Ron, each of us has a personality—a unique culture-bound consciousness embedded in autobiographical memories. Because our mental life seldom requires conscious effort, we don't even notice it. Throughout a lifetime we plan, think, fall in love, remember, forget, and become hungry, sleepy, angry, and sad. Hence, unless the biological underpinnings of human psychology are made the focus of study, as they are in this chapter, we can live our lives quite successfully, thank you, without any knowledge of how our brains function.

We take for granted that consciousness, perception, motivation, memory, and the behaviors they make possible are dependent on an intact brain. We shouldn't. The richness of the human experience begins in the brain's functioning, and psychologists are not alone in wondering about the magic of its contents. Due to specific brain structures, humans see, hear, smell, taste, touch, and feel warmth, cold, pain, and pleasure. Through the process of introspection each of us with intact brains can become conscious of and ruminate on the complexity of our mental lives. We make moral judgments. Sitting on a mall bench we may also wonder about the minds of others. (Is that person by the department store who is talking to herself rehearsing a play, doing drugs, or suffering from a mental disorder? Should I be amused or concerned?)

One benefit of studying people like Ron is that a malfunctioning brain makes us aware of the normal brain's invisible activities. Until the past few hundred years or so, we couldn't weigh happiness or memories in drams, nor poke and prod at dreams using tools other than words. But now we can. Biological psychology is the study of the mind, brain, and behavior of humans and other animals from evolutionary, behavioral genetic, physiological, and psychological perspectives. In part, biological psychology attempts to understand mind and consciousness by studying its underlying "physical substance." Neuroscientists study behavior in laboratory animals by measuring responses to brain manipulations using drugs, brain lesions, brain stimulation, and other intrusive techniques. Neuroscientists study humans as well as other animals. For example, the change in Ron's behavior following removal of a tumor from his brain is an example of a brain-behavior relationship—an understanding of the mind and behavior from the perspective of its underlying physiological and anatomical substrates (Kolb & Whishaw, 2004). Cognitive neuroscientists study perception, attention, memory, language, and conceptual processes using brain imaging, often in conjunction with other behavioral measures such as paper-and-pencil tests. An example would be to study the neural correlates of human memory using a brain-imaging technique such as *functional magnetic resonance imaging* (fMRI). (The fMRI allows researchers to image neural activity of a person while in the act—for example, reading, listening to music, learning a task.) Although each theoretical perspective of psychology provides some unique insights not offered by others, cognitive neuroscience holds special promise for an integrative understanding of brain-behavior relationships. Studying the brain activity underlying appropriate social behavior, for example, allows us to understand better how tumors could produce Ron's inappropriate social behavior.

This chapter begins by describing, from evolutionary and genetic perspectives, the overall design of the human brain and the behavior it generates. We next explore the anatomical, physiological, and pharmacological properties of individual neurons comprising the brain, and how drugs affect the neurotransmitters secreted by neurons. Last, we will return to the brain structures and functions that underlay the conscious mind.

EVOLUTIONARY AND GENETIC PERSPECTIVES

All animals, including humans, have unique mental lives and behavioral tendencies due to the size, shape, and organizational features of their brains. Chimpanzees, for example, have mental lives and behavioral tendencies more similar to humans than to chickens. On the other hand, chickens also see, hear, learn, remember, get hungry and full, and respond to pain-inducing stimuli by squawking. It turns out that the uniquely designed brains of these three species also have similar brain structures.

Why is that? The answer to this question comes from scientists who study brain-behavior relationships from different perspectives: Geneticists compare genetic recipes for proteins that comprise brains in different species, neuroanatomists and neurophysiologists, respectively, study brain structure and function of animals, and neuroscientists study brain-behavior relationships. All are guided by Darwin's evolutionary theory, which posits that brains were adapted to solve problems of survival and reproduction (Barkow, Cosmides, & Tooby, 1992; Darwin, 1859/1962). For example, an evolutionary answer to the question "Why is it human nature to sleep about 8 hours during the nighttime hours?" would be that humans who behaved in this way survived and reproduced more offspring than humans who didn't. In other words, sleeping 8 hours (as opposed to 4 hours or 12 hours) is explained as an adaptive trait of human nature. Human brains and behavior differ from chimpanzee and chicken brains and behavior because they evolved in different ecological niches, including adaptively different sleeping patterns (as well as eating patterns, social behaviors, reproductive and maternal caretaking behaviors, and so forth). But their brains and behavior are also similar because problems of survival and reproduction have common elements.

WHAT IS THE HUMAN BRAIN PLAN?

Sixty years ago, C. Judson Herrick (1948) wrote *The Brain of the Tiger Salamander*, a much-forgotten book that detailed a brain plan common to all vertebrates (fish, amphibians, reptiles, birds, and mammals). Herrick pointed out that all vertebrate species have a central nervous system (CNS) comprising a bilateral (left and right side) brain and a spinal cord. All have 12 paired cranial nerves that bring sensory information into the brain and send motor (movement) signals to facial muscles. All are sensitive to different wavelengths of light from the sun. All have eyes and "see," but not all "see" the same thing. All are sensitive to vibrations of molecules of moving air (that is, all have ears; a fish's "ears" are sensitive to vibrations in water). All have taste and smell receptors; all are sensitive to touch, including warmth, cold, and pain. These receptors all enter their respective brains in the same place—the brainstem. Moreover, all have a forebrain, a midbrain, and

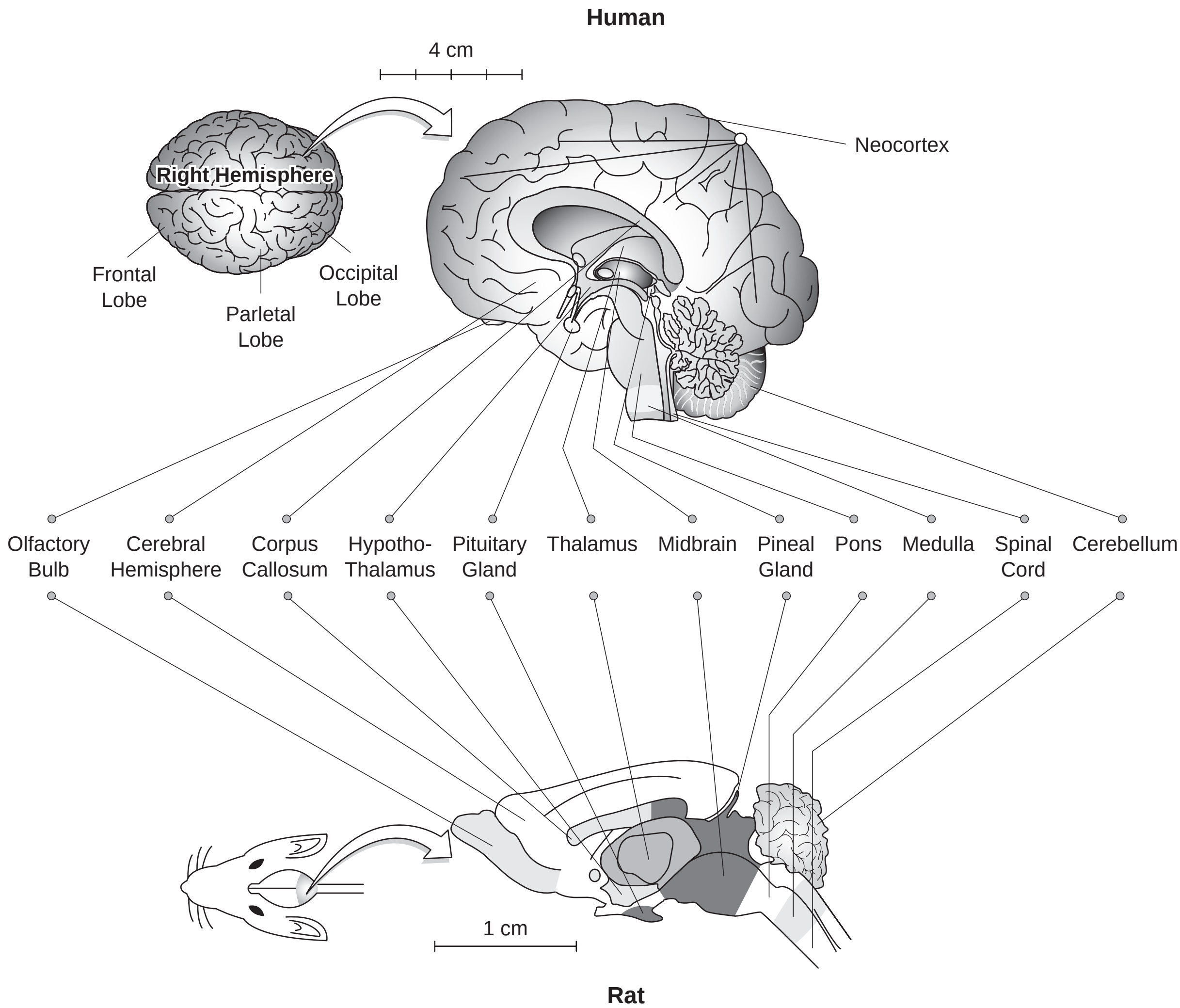

Figure 13.1 A comparison of the structures of the human brain (top) and rat brain (bottom). (Not to scale.) Disregarding size, the subcortical structures are similar, and the defining feature of the human brain is the massive overlay of cerebral cortex.

SOURCE: Modified from Figure 6.8 in Rosenzweig, Leiman, and Breedlove, 1999.

a hindbrain (see Figure 13.1)—that is, a front end, middle, and back end. (Sensory information comes into the mid-brain structures, and voluntary movement is initiated by the forebrain.)

Humans also share with vertebrates the other mid-brain structures (commonly called the brain stem), the cerebellum, hypothalamus, and thalamus. Figure 13.1 shows placement of these common structures in the human brain (top) and the rat brain.

Given all that is common, what makes the human brain so uniquely different? The short answer is the huge amount of cerebral cortex of both hemispheres that completely covers the other brain structures. The human mind and behavior, then, are the result of an ancient brain plan common to other animals that has become overlain in our recent evolutionary history by hundreds of millions of “new” neurons (*neocortex* means “new cortex”) comprising the cerebral hemispheres. The uniqueness of human consciousness lies in the functioning of this massive overlay of neurons comprising the neocortex. To understand the neocortex better, we drop down to a lower level of analysis of brain functioning—the neurons of which the brain is made.

NEURONS

Neurons are specialized cells in the brain that have abilities analogous to those of behaving animals. They are sensitive to the environment (that is, they receive information from their surroundings) and act on it. As shown on the left-hand side of Figure 13.2, each neuron has a soma, or cell body, and protrusions from it that either receive information (dendrites) or send information (the axon). One way that neurons are differentiated from most other cells in your body, then, is that a neuron is

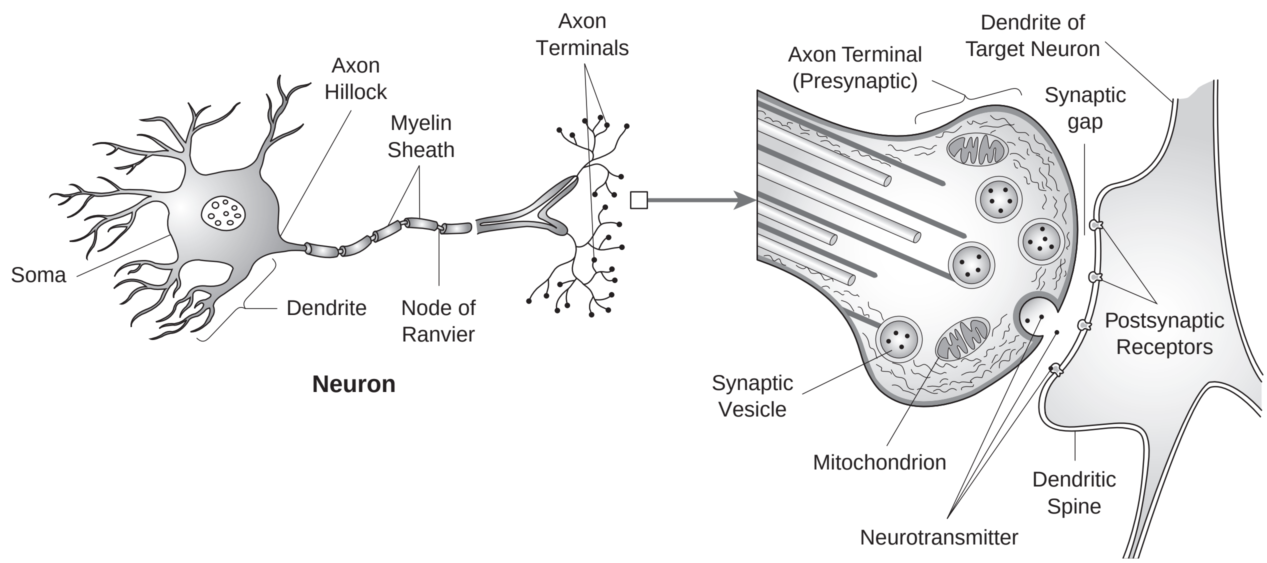

Figure 13.2 The structure of a typical neuron (left) with a blowup of the axon terminal (right), where synapse occurs.

excitable—most appropriate for cells underlying psychological activity!

Depending on how neurons are analyzed, they reveal chemical, pharmacological, electrical, magnetic, and, ultimately, psychological properties. Electrical properties are measured with voltmeters. The electroencephalogram (EEG) measures the voltage of many thousands of neurons by attaching electrodes to the scalp. MRI, *m*agnetic *r*esonance *i*maging, measures magnetic properties of thousands of neurons and by computer converts these potentials to images.

Though most function in similar ways, neurons are highly specialized. Different types of sensory neurons detect information in the environment (for example, the detection of light for seeing, vibration for hearing, and chemicals that are tasted and smelled). Motor neurons cause muscle tissue to contract, allowing us to move through the environment. (Perhaps because "moving a muscle" is conceptually simple, we tend to dismiss the role of neurons in accomplishing this basic behavioral unit. But that would be a mistake, for the abilities of skilled movements is in part what makes humans such interesting animals.)

Similarly functioning neurons in the brain are responsible for your ability to read this text, others for whether you are paying attention or nodding off. Some specialized neurons allow you to tie the first part of this sentence to the last part in what is called your short-term memory. Yet others are responsible for your impatience with this long sentence, in a long paragraph, in a long chapter. Hope to do well on your next assignment? You've got neurons responsible for your hope (or lack of it), for discipline (or lack of it), for perfecting your tennis swing, and for controlling your temper. Consciousness, perception, learning, memory, personality—all are the result of the functioning of billions of neurons in your central nervous system (Kandel, Schwartz, & Jessel, 2001).

Neurons that sense events function differently from motor neurons that activate muscles. Likewise, neurons in the cerebellum are not the same as those in the temporal lobe, and all differ from the five types of neurons in the retina of your eye. Instead of describing any one of these types, we'll first describe the functioning of a generic, unspecialized neuron. When activated, it communicates with adjacent neurons by discharging a chemical, called a *neurotransmitter*. Some neurotransmitters cause a muscle to twitch. Others are involved in seeing or in memory formation. Still others are involved in the psychological experience of pleasure and pain. Then, we'll examine how our knowledge of a neurotransmitter's function is illuminated by different psychoactive drugs. We begin by learning about the neuron's remarkable membrane "at rest," before it becomes activated.

THE NEURON'S MEMBRANE: THE KEY TO EXCITABILITY

Actually, the membrane of a neuron is never "at rest." A neuron's membrane separates its inside from its outside. The membrane of a neuron has pores, or channels, through which ions can pass. Ions such as sodium and potassium are continuously attempting to move through this membrane covering dendrites, the soma, and the axon of a neuron. These ions, or electrolytes (so called because they carry a tiny positive or negative electric charge), are not "balanced" across the membrane. Their imbalance creates an electric charge across the membrane, called a *membrane potential*. Potassium and sodium ions are both positively charged, while chloride and proteins inside the cell are negatively charged. The charge, or resting membrane potential, across the cell membrane is due to this unequal concentration of electrolytes inside the cell (in the intracellular fluid) relative to outside the cell (in the extracellular fluid). If equal amounts of positive and negative charges existed both inside and outside the cell, the resting membrane potential would be zero. The normal resting membrane potential, or voltage across a neuron's membrane before it is activated, however, is usually about –70 mV (minus 70 millevolts). (The charge is measured by sticking an electrode through the cell wall of the neuron and attaching it to a voltmeter.) The resting membrane potential is negative, not zero, because there are more negatively charged chemicals inside the cell and more positively charged ones outside. The cell is said to be at equilibrium (at rest) when there is no net movement of ions in or out of the cell.

However, at –70 mV, the neuron is "primed," or poised to discharge. Then, one of two events can occur. A slight chemical disturbance of the membrane can cause a gradual change in the resting membrane potential of the dendrite or soma of a neuron. This effect is called a *graded potential*. The voltage may slowly increase or decrease, depending on the kind of chemical disturbance on the cell membrane. However, if the graded electrical charge is sufficiently large, the entire neuron can be triggered to discharge electrically. This large nerve impulse, sometimes referred to as the firing or spiking of the neuron, is called an *action potential*. An action potential, in turn, can ultimately disturb the resting membrane potential of other neurons, affecting their excitability. In fact, the action potential is one of the most important ways by which one neuron "talks" to another.

The Action Potential

Recall that the cell membrane of a neuron, including its axon, has pores, or channels, though which ions can pass. The rapid movement of ions through the axon's membrane is the mechanism underlying an action potential. At rest, more sodium ions (Na^+) are outside, and more potassium ions (K^+) are inside the axon's membrane, electrically polarizing it. Depolarization of the membrane to its threshold level has the effect of opening both sodium and potassium channels in the membrane for about 0.5 msec. During this brief time, the Na^+ ions rush into the cell and the voltage across the membrane rises rapidly, from –70 mV, past zero volts, to about +50 mV. In response, seeking equilibrium, potassium ions (K^+) rush out of the cell through their channels, eventually reversing the membrane's polarity. This K^+ ion movement in turn triggers the adjacent Na^+ channels to close, rendering the membrane resistant to any further outflow of Na^+ ions. In another fraction of a second, the resting membrane potential is restored, and the neuron is ready to discharge again. The result is a "spike" of electricity that flows rapidly away from the cell body down the axon.

When an action potential occurs, the depolarizing effect on the cell membrane quickly spreads to adjacent areas of the axon. The result is that the action potential travels down the length of the axon (the nerve "fires"). Shortly after the action potential has passed one portion of the axon, the K^+ and Na^+ ions resume their resting membrane condition. The axon cannot "half-discharge." Once the neuron's threshold for discharge has been reached, a complete action potential follows. In contrast to graded potentials, this all-or-nothing relationship is known as the *all-or-none principle*. Following an action potential, a brief time, called a *refractory period*, must pass before another action potential can be generated. Nevertheless, a given neuron can discharge repeatedly if the neuron remains depolarized below its threshold. Some neurons can fire up to 1,000 times per second.

Most axons are literally wrapped by a different type of cell (called a Schwann cell) that does not function like a neuron. Schwann cells form a *myelin sheath* that acts as an insulator, not allowing the electrical discharge of one neuron to affect adjacent axons. Figure 13.2 (left) shows that the myelin sheath covers the entire length of the axon except for tiny gaps between the Schwann cells, called the *nodes of Ranvier*. K^+ and Na^+ ion movement can occur only at the nodes of Ranvier. The result is that the action potential jumps rapidly from node to node, a process called *saltatory conduction* (saltatory from the Latin *saltare*, meaning "to leap"). The gain in speed is substantial: The one-meter-long, large diameter, myelinated axons that run from a 7-foot basketball player's spinal cord to his toes conduct nerve impulses at about 100 meters per second. In contrast, small, unmyelinated fibers conduct impulses at a rate of only 1 to 2 meters per second. Such fibers are responsible for communicating chronic, dull aching pain.

The Release of Neurotransmitters

For present purposes we will assume that each neuron makes and releases but one type of neurotransmitter. The right-hand portion of Figure 13.2 indicates what happens when an action potential is conducted past the myelin sheath and arrives at an axon terminal. The axon terminal makes a *synaptic connection* with a target neuron or a muscle. (*Synapse* is the point at which one neuron encounters another.) Simply put, the action potential causes the release of a particular neurotransmitter into the *synaptic gap* (the space between adjacent neurons), effecting communication with another neuron. By contrast, motor neurons synapse on muscle fibers, causing them to contract.

How does this transmission of information from one neuron to the next occur? First, each neuron synthesizes its own particular neurotransmitter, which is stored in vesicles within the axon terminal. The action potential arriving at the axon terminal has the effect of opening calcium (Ca^{++}) channels, allowing Ca^{++} to enter the cell, which causes the vesicles to fuse with the membrane wall of the axon terminal. The vesicle's cell wall then opens, spilling its contents into the synaptic gap.

Should the frequency of action potentials arriving at the axon terminals increase, more Ca^{++} would enter the cell, causing more vesicles to merge with the terminal's membrane and open, dumping more neurotransmitter into the synaptic gap. Once the neurotransmitter is in the synaptic gap, its distinctively shaped molecules find *receptor sites* in the target neuron's membrane. Each neurotransmitter has a particular shape that can be thought of as a key—the postsynaptic membrane has receptor sites that can be characterized as locks into which the keys may or may not fit (Changeax, 1993).

HOW TO EXCITE A NEURON

Disruption of the membrane potential of a target neuron and the target neuron's generation of an action potential begins with the chemical events taking place at synapse. Figure 2 (right) shows how axon terminals synapse with both dendrites and the cell body of another neuron. When a neurotransmitter is released into the synaptic gap, it makes contact with the membrane of another neuron. The axon terminal is referred to as *presynaptic* (because it initiates events at synapse), and the target membrane on the other side of the synaptic gap is *postsynaptic* (because it is on the receiving end of synapse).

The contact that causes the target membrane to be excited is called an *excitatory synapse*. An excitatory synapse has the effect of decreasing, or depolarizing, the resting membrane potential of the target neuron. This depolarization is localized, meaning that it affects only a small area of the target neuron's membrane. It results in a graded potential called an *excitatory postsynaptic potential*, or EPSP.

Meanwhile, a different neurotransmitter secreted from another neuron may be having the opposite effect when it makes contact with the target neuron's membrane—it may inhibit the target neuron's membrane potential. Like excitatory neurotransmitters, inhibitory neurotransmitters affect only a small area of the target neuron's membrane. Inhibitory neurotransmitters tend to *hyperpolarize*, or increase the resting membrane potential. For this reason, hyperpolarization—an increase in the membrane potential—is called an *inhibitory postsynaptic potential*, or IPSP. Neurotransmitters that hyperpolarize the membrane, therefore, tend to work against, or to inhibit, the depolarizing effects of EPSP.

How do EPSPs and IPSPs interact? Think about EPSPs and IPSPs as sexual foreplay. Some events are excitatory, bringing you closer to achieving orgasm (the action potential); other events inhibit you. EPSPs and IPSPs are the localized, graded potentials that spread passively toward the *axon hillock* (see Figure 13.2), the site that determines each neuron's threshold of excitability (approximately –65 mV, or 5 mV below the resting membrane potential). Once this threshold is met—once the EPSP has depolarized the membrane at the axon hillock to –65mV—the neuron initiates the voltage changes known as an action potential, as previously described.

The axon hillock integrates the effects of multiple inputs from other neurons through the processes of spatial and temporal summation. *Spatial summation* (summation on the surface space) occurs when neurotransmitters from several adjacent axon terminals excite an area of the membrane near the axon hillock. *Temporal summation* (summation in time) is caused by a rapidly firing neuron's excitatory neurotransmitter depolarizing effects accumulating at the axon hillock. IPSPs, which act to block the excitatory effects on the axon hillock, can also summate spatially and temporally. The EPSPs and IPSPs algebraically summate: If the EPSP overcomes the IPSP at the axon hillock, the neuron generates an action potential.

NEUROTRANSMITTERS AND THEIR EFFECTS

To describe in detail the chemical differences of the various neurotransmitters and how they interact with drugs (a field called neuropharmacology) is beyond the scope of this chapter (see Cooper, Bloom, & Roth, 1996). Here we can provide only an overview: Table 13.1 groups some common neurotransmitters into five classes and describes their primary functions. Neurons that secrete these neurotransmitters differ both in their location in the brain and in their specific functioning.

Let us look at a few examples. The neurotransmitter dopamine is excitatory and is synthesized primarily in the brainstem of humans and other animals. Dopamine

Table 13.1 Classification of some of the neurotransmitters and their functions

Neurotransmitter	*Action*
Acetylcholine	Contracts muscle tissue; Excitatory (found in brainstem and cortex)
Glutamate	EPSP*; Excitatory in CNS
Aspartate	Excitatory in CNS
Glycine	Excitatory in CNS
GABA	IPSP** Inhibitory in CNS
Dopamine	Excitatory (brainstem and cortex)
Epinephrine (adrenaline)	Excitatory (found in brainstem and cortex)
Norepinephrine	Excitatory (found in brainstem and cortex)
Serotonin	Excitatory (found in brainstem and cortex)
Endorphins (such as beta-endorphin)	Activates neural systems

SOURCE: Modified from Pinel, 1997.
*EPSP—Excitatory postsynaptic potential
**IPSP—Inhibitory postsynaptic potential.

pathways (from dopamine-secreting neurons) extend into the cerebral cortex. In experimental work with rats, neurons secreting dopamine can be stimulated by brief pulses of low-level electricity through an implanted electrode in a structure near the hypothalamus called the medial forebrain bundle (Olds, 1969; Olds & Milner, 1954). Electrical stimulation is presumed to mimic the action potentials of neurons. Rats will press a lever to receive such stimulation in the same way that they will press a lever to receive food. These and subsequent researchers studying dopamine pathways proposed that the neurotransmitter dopamine "causes" pleasure both in rats and humans—that "pleasure" has both an anatomical and a chemical nature.

To take another example, acetylcholine (ACh) is the neurotransmitter neurons secrete onto muscle fibers and is present in other neurons in the central nervous system and autonomic nervous system. Using the lock-and-key analogy, when released, dopamine fits dopamine receptors, and acetylcholine will activate only ACh receptors.

The actions of both dopamine and acetylcholine differ from those of yet another class of neurotransmitters, amino acids, which include glutamate (see Table 13.1). Glutamate has excitatory effects, and GABA, another amino acid, is an inhibitory neurotransmitter. Other excitatory neurotransmitters include epinephrine, norepinephrine, and serotonin. A distinctive class of chemicals called neuropeptides also function as neurotransmitters. Endorphin, for example, is a neuropeptide that stimulates other neurons in the brain and spinal cord to produce analgesia (pain relief) and the experience of pleasure. Another neuropeptide, cholecystokinin, is released following a meal and produces satiety, the sensation of feeling full.

This brief summary of the neurotransmitters does not do justice to the incredible stories presently unfolding in neuroscience and neuropharmacology laboratories around the world. Sixty years ago, only two transmitters, one excitatory and one inhibitory, were known to exist. Table 1 lists just a few of many dozens of neurotransmitters that have since been discovered in humans and other animals.

Neurotransmitters made and released by neurons in discrete parts of the brain are the sources of our motivation and emotion, our voluntary behavior, the ways in which we learn and remember, and other aspects of our consciousness. Because humans share DNA with other animals and even plants, it should be no surprise that we also share some of the same neurotransmitters. For example, Hsieh, Lam, van de Loo, and Coruzzi (1998) reported finding genes in a weed, Arabidopsis, which encode for receptors sensitive to the neurotransmitter glutamate. Indeed, the number of plants that contain neurotransmitter-like chemicals that are common psychoactive drugs is amazing. Cannabis from marijuana and poppies that produce opium and tobacco leaves containing nicotine are well-known examples. Let us look at how some of these drugs affect the normal functioning of neurotransmitters.

DRUGS, NEURONS, AND BEHAVIOR

Table 13.2 summarizes how drugs affect human consciousness through their action on neurons. Drug categories include depressants (natural and human-made), antianxiety drugs (human-made), stimulants (natural and human-made), antidepressants (human-made), narcotics (natural and human-made), and psychedelics and hallucinogens (natural and human-made).

Depressants such as ethanol (beer, wine, and spirits) alter waking consciousness by depressing it, or slowing it down, by inducing relaxation and reducing physical activity. Alcohol stimulates the release of dopamine. Dopamine is also released by other drugs, such as cocaine, and, as

Table 13.2 Classification of major psychoactive drugs

Drug Classification	*Common Name*	*Drug Effect*
Depressants		
Alcohol	beer, wine, spirits	relaxation, buzz, mild euphoria
Barbiturates (tranquilizers)	Nembutal, Seconal	mellow; mild sedation
Antianxiety Drugs		
Benzodiazepines	Valium, Ativan, Xanax	calming, anxiety reduction
"second generation" anxiolytic	BuSpar	
Stimulants		
Amphetamines	methamphetamine, Ritalin, Dexedrine	stimulate, rush, increase alertness (also elation and nervousness, dose dependent)
Cocaine	coke, crack	
Nicotine	tobacco	
Caffeine	coffee, tea	
Antidepressants		
Tricyclics	Imipramine, Elavil	relieve depression, normalize mood, uplift
SSRIs	Prozac, Effexor, Zoloft	
Narcotics		
Opiates	morphine, heroin	mellow, stupor, euphoria
Psychodelics and Hallucinogens	Marijuana, hashish, LSD	alter time and sensory perception, distort consciousness

SOURCE: Modified from Pinel, 1997.

described earlier, by electrical stimulation of the brain, by the ingestion of tasty food, or by reading a good book. Alcohol also alters consciousness by activating GABA receptors, depressing neural activity in the frontal cortex, which normally inhibits behavior. Thus, a moderate dose of alcohol disinhibits inhibited behavior; a person may therefore exhibit heightened sexuality, aggressiveness, or playfulness. These drug effects are dose-related; higher doses produce slower sensory-motor reaction times, disordered thought, slurred speech, poorer-than-normal judgment, and poor coordination. At the highest doses, alcohol becomes extremely dangerous: Sleepiness, unconsciousness, coma, and death can occur quickly as neurons in the brainstem that regulate consciousness are disabled.

Caffeine is a stimulant, a sympathomimetic (mimics the action of the sympathetic nervous system—the "fight or flight" system of arousal). The sympathetic division of the autonomic nervous system is aroused due to caffeine-stimulated release of the neurotransmitter epinephrine from adrenal glands. The result: dilation of the lungs' bronchi, constricted blood flow to the brain, and enhanced blood flow to the rest of body. Another stimulant, amphetamine, overactivates dopaminergic and serotonergic neurons. Chronic use of high doses of amphetamine can cause permanent damage to the neurons that release dopamine and serotonin, producing "amphetamine psychosis"—severe paranoid hallucinations associated with aggressive, destructive behavior, as well as sleep disturbances, sexual dysfunction, depression, poor memory, loss of coordination, and even schizophrenia (Julien, 1995).

The primary effect of the stimulant cocaine is to block the reuptake of dopamine, allowing the neurotransmitter to remain longer in the reward and pleasure centers near the hypothalamus (see Figure 13.1). Nicotine produces its pleasurable effect by releasing epinephrine from various sites in the body. (Epinephrine, among other effects, causes CNS arousal.) It also releases an endogenous opiate, beta-endorphin. (*Endogenous* means "normally occurring.") Because nicotine also stimulates acetylcholine receptors and dopamine receptors, tolerance to nicotine develops rapidly, making it one of the most highly addictive drugs.

Opiates (narcotics) such as morphine and heroin alter consciousness by producing analgesia and pain relief and, at higher doses, stupor and euphoria. These drugs' "keys" perfectly fit the "locks" of opiate receptors, receptors that are also stimulated by naturally occurring "endogenous opiates" called *endorphins*. (The term *endorphin* comes from "endogenous morphine.") Endorphins are released naturally when physiological systems are stressed. An example is endorphin release during and following an intense workout such as a competitive athletic event. The feeling of well-being during recovery is caused by endorphin release.

Hallucinogens such as marijuana, hashish, and LSD are all quite different compounds. Their chemical structure can be similar to serotonin, norepinephrine, and acetylcholine. How the various hallucinogens work is less well understood. For example, LSD and the designer drug ecstasy (*X*) are serotonin antagonists—they block serotonin receptors. The cannabis in marijuana fits perfectly in cannabinoid receptors found throughout the brain (the nucleus accumbens in the brainstem, cerebral cortex, hippocampus, cerebellum, basal ganglia, hypothalamus, brainstem, and spinal cord—even the spleen!). These cannabinoid receptors function as *neuromodulators*, altering the function of norepinephrine, dopamine, serotonin, acetylcholine, and other neurotransmitters.

From these examples one can appreciate how the biology of the brain provides a better understanding of mind and behavior. Normal brain functioning comprises myriad interactions of billions of interconnected neurons secreting functionally different neurotransmitters. Indeed, the model of EPSPs and IPSPs summating on a single neuron presented above does not begin to capture the complexities of brain functioning. Each neuron may have 10,000 or more places of synapse upon it, integrating information from thousands of different neurons.

One last example focuses on the neurotransmitter serotonin. Human-made drugs with the brand names Prozac, Zoloft, Welbutrin, and Effexor affect how much serotonin is in the synaptic gap between two neurons. In people who are neither too sad nor too happy, an "optimal" amount is released by serotonergic neurons. If too little serotonin is present, a person feels sad and depressed. By adding an SSRI, a *selective-serotonin reuptake inhibitor* (Prozac is an example), the amount of serotonin in the synaptic gap increases, and symptoms of depression are diminished.

IS HUMAN CONSCIOUSNESS MERELY BRAIN CHEMISTRY?

That drugs interact with neurons to affect our conscious experiences raises other interesting questions. Are our mental and emotional lives merely chemical interactions? This question has not been (and may never be) answered by neuroscience researchers. In contrast to the underlying chemistry of consciousness, humans *experience* the phenomena of consciousness in what seem to be qualitatively different ways. The pain of a bee sting is so unlike the memory of last night's dream, and our experience of each seems so unlike the chemical changes of neurons. Humphrey (1992) proposed that human consciousness is an emergent property of the interactions of billions of neurons in the large human brain. (Emergence is the unforeseen result of the synthesis of individual elements.) But not everyone thinks that brain chemicals provide a satisfying "explanation" of human (Chalmers, 2007) and nonhuman, animal consciousness (Wasserman, 1993). Certainly, "chemical, pharmacological, electrical, and magnetic" are forms of energy, but as yet, seeing, hearing, and tasting are not translatable into any of them. Therefore, it is reasonable to conclude that knowledge of the brain doesn't "solve" the mind-body problem; rather, it is a highly promising approach.

THE HUMAN BRAIN

So far we have explored the evolutionarily and genetically determined brain plan of vertebrates, including humans. The overall design, including bilateral organization, spinal cord, brainstem, cerebellum, thalamus, hypothalamus, and cranial nerves, is common to all vertebrates. This early brain plan was adaptively selected by the environment in which life evolved—one that was sensitive to sunlight; vibrations in water (and later, in the earth's atmosphere); chemicals (food and other animals) to be tasted and smelled; and motor systems to move, capture prey, avoid predators, find mates, and so forth. We have also examined the properties of the basic functioning units of these vertebrate brains, their neurons, and how drugs affect their neurotransmitters. Next we explore further how the human brain differs from the brains of mice, monkeys, and chimpanzees to afford us our unique consciousness and behavior.

Compare again the rat brain with the human brain in Figure 13.1. Note they are not drawn to scale. The mouse brain is much smaller than a pencil's eraser. You can approximate the size of your own brain by putting your two closed fists together at the knuckles, each fist representing a left and right cerebral hemisphere. But the comparison is not fair, you might argue, because the human body is far larger than a rat's is, so of course our brain is larger. You would be right in that there is a brain-mass/body-mass ratio that can be computed that would correct for body size. Even so, in such a ratio, humans remain "over-brained." All the great apes (including humans, chimps, and gorillas) fare well in such brain-body ratio comparisons, but humans are the most outstanding (Jerison, 1973). Furthermore (and this is the key to human uniqueness), most "extra" neurons are in the outermost covering of the brain, the cerebral cortex. The subcortical (below the cortex) brain of a chimpanzee is extremely similar to that of a human. The chimpanzee brain is also covered by the neocortex (new cortex), which makes up the cerebral hemispheres. But for humans, the surface area of the neocortex, comprising billions of neurons, far exceeds that of a chimpanzee.

The wrinkled appearance of a human brain, with its hills (gyri) and valleys (sulci), is deceptively small. If you were to "skin" the cerebral cortex off the subcortical areas it completely covers, unfold it, and spread it out, it would be about the size of a large road map. (In demonstrating this point in lectures, I bring in a road map squashed into a baseball-sized ball, and unfold it.) This huge amount of "new" cortex developed fairly recently (0.5-1 mya) in our evolutionary history. It is this explosion of brain tissue that differentiates us from our hooting, cultureless, primate relatives. This neocortex and its interconnectedness with an immediately underlying structure called the thalamus (see Figure 13.1) is what makes possible human consciousness, including thought, language, skilled movement, and culture (Edelman & Tononi, 2000). Why can't a chimpanzee be taught language or sing a song? It doesn't have a sufficient amount of neocortex or the appropriate organization of its limited neocortex.

This neocortex comprising the left and right cerebral hemispheres is conventionally divided into four lobes: frontal, parietal, temporal, and occipital (see Figure 13.1). Think of these "lobes" as rough anatomical and functional guidelines. For example, most of the occipital lobe (at the back of the brain) is made up of neurons involved in various activities of "seeing," and neurons in a small portion of the left temporal lobe (on the left side of the brain at ear level) are necessary for "hearing and understanding language." This approach emphasizes the brain's specificity of function—each part of the brain is involved in a discrete functional activity.

The Brain's Plasticity

Alternatively, the neocortex covering the human brain also exhibits what is called "plasticity." The neocortex can acquire new functions through experience with the environment. To take an extreme example, if the so-called "language neurons" on the left side of the brain are damaged in a small child, the child no longer understands spoken language. Amazingly, with time, a commensurate area in the *right* temporal lobe is somehow induced to assume this function and the child can relearn language. (Unfortunately, this capability is mostly restricted to juvenile brains; very little language is relearnable if the damage occurs after puberty.)

Another example of the brain's plasticity is that neurons "reorganize" themselves during a lifetime depending on what behavior a person engages in—a literal "use it or lose it" situation. Into music for a couple of years? Your brain organization will reflect this activity, and will change to reflect new activities in new environments such as playing chess for an extended number of years. A final example of the brain's plasticity is called *neurogenesis* (Eriksson et al., 1998). New neurons appear in a part of the brain's neocortex called the hippocampus when you encode large amounts of information into memory.

A human's subcortical structures (common to humans, mice, and monkeys) reflect more specificity of function and less plasticity. Table 13.3 makes this point: The behaviors common to humans and other animals in the left column, though not all subcortically mediated, are more "hard-wired." The reason is that the brain of an animal is first and foremost designed to keep it alive. The subcortical regions appeared millions of years before the cerebral hemispheres of the human brain evolved, and "plasticity of behavior" certainly doesn't characterize how fish, frogs, and snakes live their lives!

This ancient brain design is reflected in how humans sense the world. The body's skin senses begin with receptor neurons that transmit information about touch, pain, and temperature. Their axons come into the spinal cord, ascend along pathways through the midbrain, and synapse in the thalamus (see Figure 13.1). Taste and sound

Table 13.3 Behaviors common to all animals (left) and unique to humans (right)

All Mammals	*In Addition, Humans*
Perceive, learn, think	Talk, read, write
Remember, eat, drink	Compose symphonies
Sleep, dream	Build bridges and computers
Date (court), mate	Play soccer, make wine
Care for young	Appreciate the universe
Engage in social behaviors	Share consciousness with each other
Common brain structures and functions	Uniquely human brain structures and functions

receptors send their axons into the medulla, a brainstem structure, where they synapse and ascend to the thalamus. Receptors in the eye are connected to nerves that project to the thalamus. The point is, sensing the world "begins" in the thalamus, a phylogenetically older brain structure, before being "completed" by neurons in the newer cerebral hemispheres.

To appreciate better the connectivity between the old and new brain in humans, we'll look more closely at how we see. Four specialized neurons in the retina of the eye are sensitive to light. In dim light, the *rod* receptors signal the presence or absence of light. The remaining three visual receptors, called *cones*, require a relatively high intensity light, and each is maximally sensitive to the wavelengths comprising the visible spectrum that are labeled red, yellow-green, and blue. When stimulated, these receptors synapse on *bipolar cells*, which in turn synapse on *ganglion cells*—both are also specialized retinal neurons. Their activity is modulated by two other types of neurons in the retina, called *horizontal cells* and *amacrine* cells. This 5-layer retina begins to organize the visual information, but "seeing" is accomplished further in the brain.

The axons of the ganglion cells leave the eye and form the optic nerve. Most of the fibers involved in the conscious experience of sight project to the thalamus. Other optic nerve fibers go to the brainstem, where they synapse on neurons that in turn reflexively control "mechanical" aspects of the eye such as dilation and constriction of the pupil (to regulate how much light enters the eye). Some optic nerve fibers project to the hypothalamus and are involved in regulating our *circadian rhythms*, including alertness, sleep, and metabolic functions (our body is warmer when light enters the eye and colder in the dark). Approximately 150,000 "non-seeing" optic nerve fibers project to another brainstem structure called the *superior colliculus*. The superior colliculus's function is to turn our head reflexively, unconsciously, to direct our gaze. This portion of the brain allows us to drive our cars "unconsciously" by coordinating what we see with how we steer, paying little attention to the process.

Conscious sight is accomplished by millions of optic nerves that synapse first in the thalamus and then by new fibers from the thalamus to *striate cortex*, the *primary visual sensory* area located in the occipital lobe. Seeing can be interrupted at any point. For example, people with damaged portions of striate cortex have blind spots in their visual field, even if the retina and thalamus are intact.

What is seen? And why is 40 percent of a human's neocortex—many hundreds of millions of neurons—devoted to "seeing"? Why can't the millions of neurons of the subcortical areas get this job done? Answers to these questions have come from neuroscientists studying the connections of neurons in striate cortex to what is called *extrastriate* cortex, or *secondary* visual sensory areas, located in the occipital lobe, parietal lobe, and temporal lobe. The picture that emerges is that the secondary (and tertiary) areas of extrastriate cortex integrate information arriving from two separate retinas into a unified percept of color, form, and movement—our visual world.

But our language reflects that "seeing" is more than the percept of form and color. "I'll believe it when I see it with my own two eyes!" "Oh, I see!" So, "seeing" is also psychologically believing and understanding. In his book, *The Man Who Mistook His Wife for a Hat*, neurologist and writer Oliver Sacks (1987) described a patient who suffered damage to one area of extrastriate cortex. He could see just fine, but he didn't know and didn't understand what his eyes were telling him. He could see his hat, and he could see his wife's head, but he confused the two! He suffered a *visual agnosia*—seeing, but not understanding what he was seeing.

WHAT IS THE FUNCTION OF THE NEOCORTEX?

The lesson taught by such brain damage gives insight into the functions of massive numbers of interconnected neurons comprising the four lobes of the neocortex. Visual generated by thalamus and striate cortex are meaningless by themselves. Their meaning derives from integrating them with prior images remembered over a lifetime. A face is familiar or not depending on a memory of having seen the face before. Extrastriate cortical neurons encode such memories. A glass is only a glass if one "knows" what a glass should look like. It has functions as well as visual features. It has a "feel" when picked up, makes a clinking sound during toasts, and produces a distinctive "taste" when wine is drunk from it. If the wine is sour, it violates our expectations. Expectations are the neurons that integrate the information from visual, tactile, auditory, and taste neurons and compare it with what has been previously experienced.

Much of the parietal, temporal, and frontal lobes are comprised of neurons that have been "trained" through experience to integrate our separate sensory experiences into a unified consciousness. The neocortex accomplishes much more than integration, though. Humans are able to "read" emotion by simply glancing at a face. (Women are better at this task than men!) Such ability is mediated by neurons in the frontal and parietal lobes that integrate "emotional information" (fear, rage, love) generated by ancient subcortical areas of the brain with visual information of "seeing" a face. Cortical brain damage diminishes this ability (Bowers, Bauer, Coslett, & Heilman, 1985). The frontal lobes can also be surgically disconnected from subcortical areas by a procedure called a prefrontal lobotomy, disrupting the individual's conscious experience of emotion.

SUMMARY

The human brain can be thought of as the integration of an "old" brain with a "new" brain. The thin veneer of neocortex covering the more primitive structures of the human brain is what allows us, unique among animals, to be civilized. Both old and new brain parts comprise neurons that function to sense one another's neurotransmitters as well as sense the outside world. Neurons also connect brain to muscle, allowing highly skilled voluntary movement. Specialized neurons in the neocortex receive lifelong training, a social process called *acculturation*. These cells' integrative functioning allows each of us a unique personality comprising hopes, plans, and expectations.

Let us conclude this brief tour of the biological underpinning of our psychological lives by revisiting Ron, whose personality changed when a tumor disrupted the normal functioning of his orbitofrontal cortex (part of the frontal lobes located just above the orbit of the eye). The neurons there integrate information from the oldest parts of the brain—emotion associated with sociality and sexuality—and from all other parts of the brain. Ron's tumor interfered with neurons that had afforded him impulse control and good judgment acquired through a lifetime. These "civilizing" neurons became dysfunctional; both old- and new-brain structures ran amuck, generating antisocial behavior. It is left for philosophers and theologians to ponder whether his behavior was "immoral."

REFERENCES AND FURTHER READINGS

Barker, L. (2004). *Psychology*. Upper Saddle River, NJ: Pearson Custom Publishing.

Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Bowers, D., Bauer, R. M., Coslett, H. B., & Heilman, K. M. (1985). Processing of face by patients with unilateral hemispheric lesions. I. Dissociations between judgments of facial affect and facial identity. *Brain Cognition, 4*, 258–272.

Chalmers, D. (2007). The hard problem of consciousness. In M. Velman & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 225–235). Oxford: Blackwell.

Changeax, J. P. (1993). Chemical signaling in the brain. *Scientific American, 269*, 58–62.

Choi, C. (2002, October 21). Brain tumour causes uncontrollable paedophilia. *NewScientist.com* news service. Retrieved June 7, 2007, from http://www.newscientist.com/article/dn2943-brain-tumour-causes-uncontrollable-paedophilia.html

Cooper, J. R., Bloom, F. E., & Roth, R. H. (1996). *The biochemical basis of neuropharmacology*. New York: Oxford University Press.

Darwin, C. (1962). *On the origin of species by means of natural selection*. New York: Collier Books. (Original work published 1859)

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. Boston: Basic Books.

Eriksson, P. S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A. M., Nordborg, C., Peterson, D. A., et al. (1998). Neurogenesis in the adult human hippocampus. *Natural Medicine, 11*, 1313–1317.

Herrick, C. J. (1948). *The brain of the tiger salamander.* Chicago: University of Chicago Press.

Hsieh, M. H., Lam, H. M., van de Loo, F. J., & Coruzzi, G. (1998). A PII-like protein in Arabidopsis: Putative role in nitrogen sensing. *Proceedings of the National Academy of Sciences, 95*(23), 13965–13970.

Humphrey, N. (1992). *A history of the mind: Evolution and the birth of consciousness*. New York: HarperPerennial.

Jerison, H. J. (1973). *Evolution of the brain and intelligence*. New York: Academic Press.

Julien, R. M. (1995). *A primer of drug action* (7th ed.). New York: Freeman.

Kandel, E. R., Schwartz, J. H., & Jessel, T. M. (2001). *Principles of neural science* (4th ed.). New York: McGraw-Hill.

Kolb, B., & Whishaw, I. Q. (2004). *Fundamentals of human neuropsychology* (6th ed.). New York: Freeman.

Olds, J. (1969). The central nervous system and the reinforcement of behavior. *American Psychologist, 24*, 114–118.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain. *Journal of Comparative & Physiological Psychology, 47*, 419–427.

Pinel, J. (1997) *Biopsychology* (4th ed.). Boston: Allyn & Bacon.

Rosenzweig, M. R., Leiman, A. L., & Breedlove, S. M. (1999). *Biological psychology*. Sunderland, MA: Sinauer.

Sacks, O. (1987). *The man who mistook his wife for a hat*. New York: HarperPerennial.

Wasserman, E. A. (1993). Comparative cognition: Beginning the second century of the study of animal intelligence. *Psychological Bulletin, 113*, 211–228.

14

NEUROTRANSMISSION

SHARON PEARCEY

Kennesaw State University

Neurotransmission involves the sharing of information within the nervous system from one neuron, or nerve cell, to another membrane (i.e., another neuron, muscle, or gland). This process entails a passage of electrical and chemical information from one neuron to many other neurons or organs. In order to explain the process of neurotransmission, we must first understand the key components of the neuron, the characteristics of the neuronal membrane, the process of stimulating the cell, the action potential, and neurotransmitter release.

PARTS OF A NEURON

The human body contains approximately 100 billion neurons with cell bodies ranging in size from 6–8 µm to 80 µm in diameter (Kandel, 1991). Researchers have divided neurons into three main categories based on function: sensory neurons, motor neurons, and interneurons. *Sensory neurons*, or afferents, transmit sensory information from the sense organs (i.e., eyes, skin, etc.) *to* the central nervous system (CNS). Contrastingly, *motor neurons*, or efferents, send information *away* from the CNS to the muscles and glands of the body. *Interneurons* are tiny cells that are found only in the CNS. They only communicate information within their cell clusters (Katz, 1966).

Although there are many types of nerve cells, there are basic components of a neuron that differentiate it from other types of cells. A typical neuron has three main parts: the receiving, the transmitting, and the communicating components (see Figure 14.1). The receiving part of the cell consists of the cell body (soma) and dendrites. Typical to most animal cells, the cell body houses the structures needed for the maintenance of the cell such as the nucleus, the mitochondria, and the ribosomes. The nucleus is considered the “brain” of the cell and contains all of the genetic material (i.e., DNA and chromosomes) for cell maintenance. Metabolic activities for the cell occur in the mitochondria, the “powerhouse” of the cell. Mitochondria produce adenosine triphosphate (ATP), the fuel source for cellular activities (Thomas, 1989). Ribosomes are small structures that translate genetic information into proteins that are used both inside and outside of the cell.

Dendrites, also a receiving component of the cell, are branch-like structures that protrude from the cell body. Smaller appendages that form on the dendrites are called dendritic spines. Dendritic spines are instrumental in the communication between neurons by receiving many input signals from other cells. In fact, primates housed in enriched environments have an increased number of dendritic spines compared to those who were not housed in enriched environments (Kozorovitskiy et al., 2005). The dendrites and cell body of only one neuron can receive information from over 50,000 axon terminals from other cells.

The axon is a tubular structure ranging from .2 to 20 µm diameter and up to 1 m in length (Kandel, 1991); it is the transmitting part of the cell. Axons exit the cell body

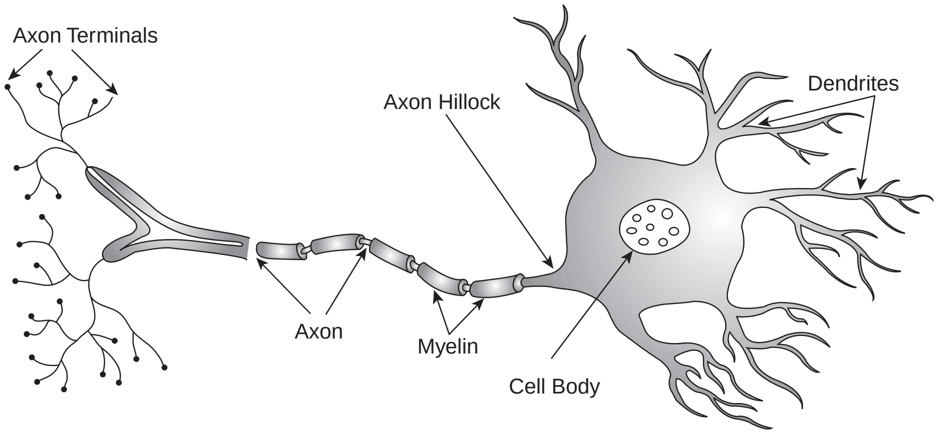

Figure 14.1 Typical motor neuron.

at the point called the *axon hillock* (Figure 14.1). In many neurons the axon is insulated by segments of protein and fat called *myelin*. Myelin is formed by Schwann cells in the peripheral nervous system (PNS) and oligodendrocytes in the CNS. Although the function of the myelin is the same in both nervous systems, increasing the speed of the signal down the axon, the manner in which Schwann cells and oligodendrocytes form myelin is very different. In the PNS, each Schwann cell wraps around the axon to form one segment of myelin. In the CNS, oligodendrocytes develop arms that wrap around nearby axons, forming myelin segments. Each oligodendrocyte can form several myelin segments on axons of many different cells. Oligodendrocytes also provide structure and support for neurons in the CNS. Each segment of myelin is separated by tiny gaps in the insulation called Nodes of Ranvier. These breaks in insulation are very important for the propagation of the action potential down the axon (Katz, 1966; see section on action potential for additional details).

The communicating part of the neuron consists of the axon terminals. As the axon reaches the end of the neuron, it branches into smaller, thin appendages with protuberances at each end (*axon terminals* or *synaptic buttons).* Within the synaptic buttons are small spherical structures called *synaptic vesicles* that house the neurotransmitters until they are released (see Figure 14.1).

TYPES OF NEURONS

Three main types of neurons can be classified by structure: monopolar neurons, bipolar neurons, and multipolar neurons. Monopolar neurons, also called unipolar neurons, have only one process that exits the cell body. These cells are most often found in the spinal cord and PNS. Bipolar neurons have one dendrite and one axon exiting the cell body and are primarily found in the PNS. The majority of neurons in the central and peripheral nervous systems are multipolar; they have an axon and many dendrites. The Purkinje cell of the cerebellum and the pyramidal cells, named for their triangular shape and found in several layers of the cerebrum, are just two examples of multipolar neurons (Figure 14.2).

CHARACTERISTICS OF NEURONAL MEMBRANES

Most of what we know about the characteristics of neuronal membranes, also called plasma membranes, has come from the experiments using the squid giant axon (Hodgkin & Katz, 1949). The large nerve of the squid (*Loligo pealli*) has an axonal-like structure that spans the animal's length (F. O. Schmitt & O. H. Schmitt, 1940). This nerve serves as the squid's CNS and has a similar membrane structure to those found in human cells. Although still very small, .5 μm in diameter, the axon is large enough to measure ionic changes across the membrane.

The plasma membrane is composed of phosphoglyceride molecules, each with a phosphoric acid head and glyceride tails made of fatty acids and hydrocarbons (Thompson, 1993). The lipid segments of the membrane are hydrophobic (repel water) whereas the phosphates are hydrophilic (seek water). These opposing characteristics create a two-layered *phospholipid bilayer* where the lipid molecules are attracted to each other, shielding it from the fluid inside and outside of the cell.

The phospholipid bilayer is approximately 6–8 nm thick (Kandel, 1991) and is selectively permeable, letting some particles flow freely through the membrane while blocking others from passing (Bean, Shepard, & Chan, 1968). Some particles need specialized protein channels to travel through the membrane (this will be discussed later).

The Resting Membrane Potential

There are many features of the plasma membrane and surrounding fluids, both inside (intracellular) and outside (extracellular) of the cell, that contribute to the resting membrane potential (Hodgkin & Huxley, 1945, 1952). A potential is calculated by measuring the difference in charge between two points. For a neuron, this resting potential is defined by the difference in electrical current between the charge inside the cell and the charge outside of the cell. At rest, a neuron's potential is approximately –70 millivolts (mV) in that the inside of the cell is more negative than the outside of the cell. This potential is created by the distribution of ions (charged particles) across the

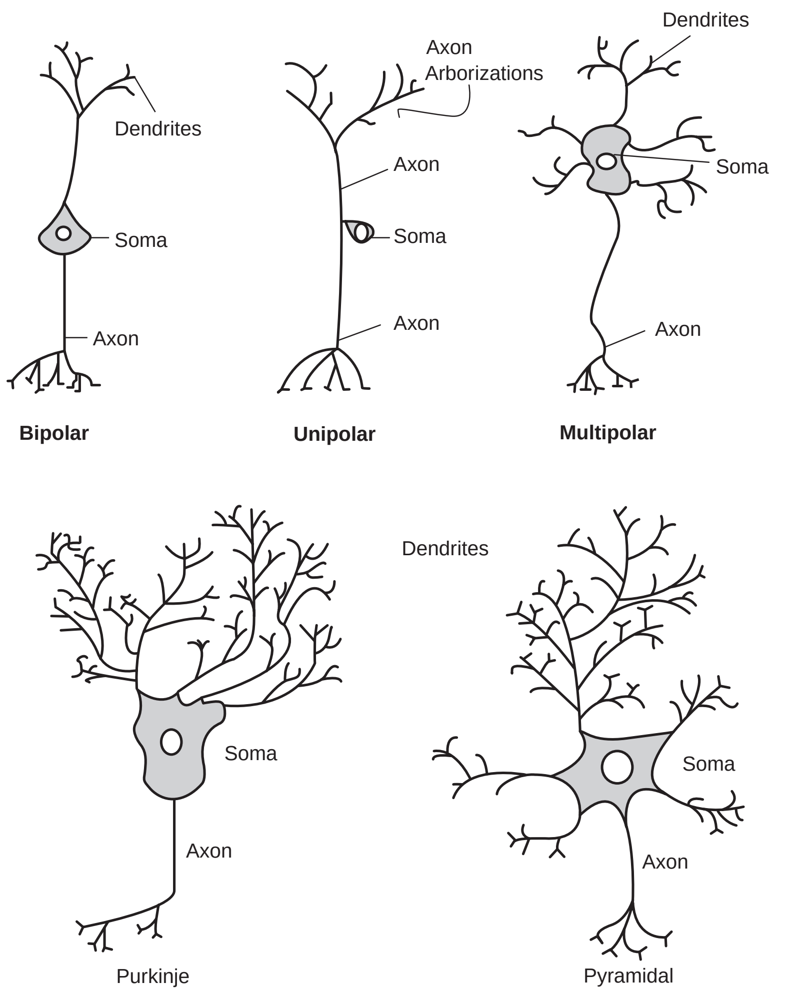

Figure 14.2 Several types of neurons.

SOURCE: Soderquist, David R. (2002). *Sensory Processes.* Thousand Oaks, CA: Sage.

membrane. Ions can be either positively charged (*anions*) or negatively charged (*cations*). The ions that create the resting membrane potential are sodium (Na+), potassium (K+), chloride (Cl–), and large organic anions (A–). There are much higher concentrations of Na+ and Cl– outside of the cell compared to the inside of the cell. Contrastingly, the inside of the cell has a higher concentration of K+ and large negative organic anions (A–) that cannot cross the membrane (see Table 14.1; adapted from Koester, 1991). These concentrations are maintained by characteristics of the membrane and forces that act on the ion distribution.

Forces Involved in the Resting Membrane Potential

Three forces involved in the maintenance of the resting potential are electrostatic pressure, diffusion, and the sodium potassium pump. Electrostatic pressure and diffusion are passive forces, requiring no energy from the cell. Electrostatic pressure involves the idea that particles of opposite charges are attracted to each other. For example, if you put Na+ and Cl– in a medium, perhaps water, the Na+ particles would be attracted to and bind to the Cl– ions, forming sodium chloride or table salt. Diffusion is a similar concept in that like particles will diffuse somewhat evenly throughout a medium. The nature of diffusion is based on the concept of electrostatic pressure in that like particles have the same charge and, thus, repel each other.

The third force that regulates the resting membrane potential is the Na+/K+ pump. This structure is a protein channel that actively pumps Na+ out of the cell and brings K+ back into the cell. For every three Na+ ions pumped out of the cell, two K+ ions are pumped back into the cell. These channels are located on all parts of the plasma membrane and require a significant amount of energy (ATP) from the cell (Rakowski, Gadsby, & De Weer, 1989; Sagar & Rakowski, 1994).

Ion Concentrations in the Resting Membrane Potential

At rest, there is a high concentration of K+ inside of the cell. The force of diffusion acts to push it out of the cell; however, the outside of the cell is more positive than the inside. Thus the force of electrostatic pressure helps keep K+ inside the cell. A similar situation occurs with Cl–. Cl– is in a high concentration outside of the cell with diffusion pushing it in; however, the negative charge on the inside of the cell works to keep it out of the cell. This is not the case with Na+. Na+ is in abundance outside of the cell and both diffusion and electrostatic pressure (Na+ is positive and the inside of the cell is negative compared to the outside) are drawing it to the inside of the cell. In order to keep the resting membrane potential at –70 mV, the Na+/K+ pump is needed to pump out Na+ and bring in K+ (see Figure 14.3).

Table 14.1 Typical concentration of ions inside and outside of the neuronal membrane

Ion	*Intracellular Fluid (mM)*	*Extracellular Fluid (mM)*
Na+	50	440
K+	400	20
Cl-	52	560
A-	385	None

SOURCE: Adapted from Koester, 1991.

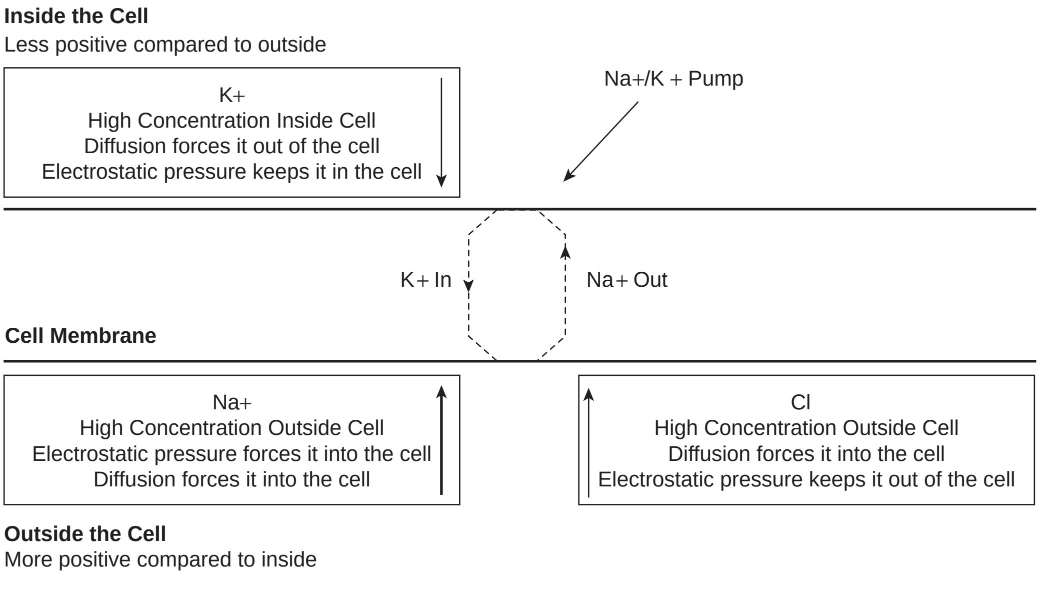

Figure 14.3 Ion concentrations and forces during the cell at rest.

Ion Channels

Although the plasma membrane is permeable to certain molecules, specialized protein channels are needed for many ions to pass through the membrane. These channels are in abundance throughout the membrane on all parts of the cell. Simply, the channels are proteins that are embedded through the membrane, acting as conduits for passing ions across the membrane. These pores are selective to certain ions, some letting only one particular ion through and others letting many positive ions pass. The selective nature of the channels is related to the size and charge of the ion, and the hydration of the ion passing through the membrane (Hille, 1967, 1971). Protein channels can be nongated or gated. Nongated channels are always open. For example, K+ is 20 times more permeable to the membrane because it is continuously leaking out of the cell through a nongated protein channel (Kandel & Siegelbaum, 1991).

Gated channels must have a mechanism that opens them, letting certain ions in or out of the cell. When the channel is open, ions can pass freely through the membrane. Contrastingly, when it is closed, ions cannot pass through. Gated channels can be ligand-gated, voltage-gated, or concentration-gated.

Ligand-gated channels have a receptor cite on the outside segment of the channel. A molecule, such as a neurotransmitter, must bind to this receptor site in order for the channel to open. Ionotropic channels are one type of ligand-gated channels. When a ligand binds to the channel, the channel opens, letting ions pass through the membrane. Ionotropic channels are fast-acting, completing ion transport in less than a millisecond. Metabotropic channels also have binding sites; however, their mechanism of action is more complicated. The channel consists of a long strand of protein that winds through the membrane seven times. When a ligand binds to the receptor site, the channel goes through a conformational change, bending and activating a second messenger inside of the cell (Kandel & Swartz, 1991). This activation causes a cascade of changes within the cell including indirectly opening ion channels from the inside of the cell and causing long-term genetic changes within the cell.

Voltage-gated channels have gates that open when the intracellular fluid has become more positive and the cell is depolarized. Upon opening, ions can cross the membrane. The gates are only open for ms and then they close, blocking the passage of ions. These channels are essential for the action potential to occur. Concentration-gated channels work on the premise of concentration gradients. When the concentration of a particular ion, either inside or outside of the cell, gets high enough, concentration-gated channels open and let ions pass across the gradient (from high concentrations to lower concentrations).

COMMUNICATION BETWEEN NEURONS

The synapse consists of the axon terminal of one cell and the cell body or dendrite of another cell. The sending neuron is called the *presynaptic cell* and the receiving neuron is called the *postsynaptic cell*. The axon terminal does not make physical contact with the postsynaptic cell; rather, there is a small gap between the presynaptic and postsynaptic cells called the *synaptic gap* (Figure 14.4). When the action potential occurs in the presynaptic cell, a neurotransmitter is released from the axon terminal. A list of some common neurotransmitters and their effects can be seen in Table 14.2 (Julien, 2001). The neurotransmitter then binds to the receptor sites on the ligand-gated channels, letting ions pass through the cell membrane. The flow of ions through the membrane causes a change in the electric potential of the cell. This alteration in charge is called a postsynaptic potential because the alteration occurs in the receiving, or postsynaptic, cell.

Postsynaptic potentials can be either excitatory or inhibitory. An excitatory postsynaptic potential (EPSP) occurs when an ion channel opens and Na+ enters the cell. The positive charge of this ion brings the cell's potential

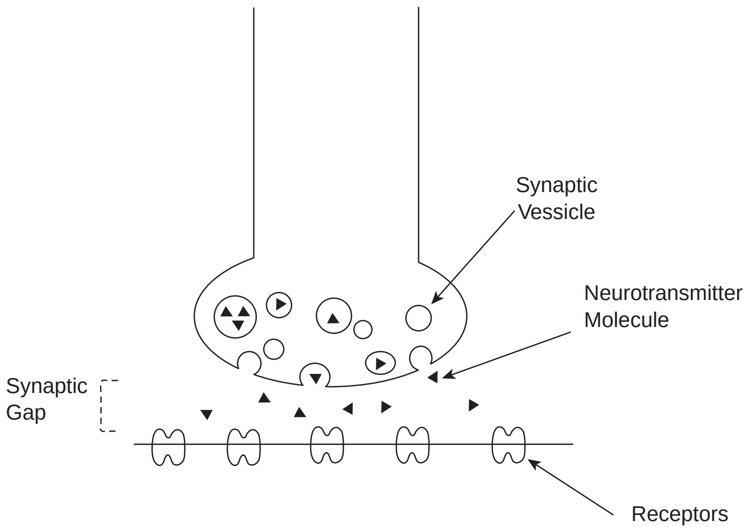

Figure 14.4 Synapse.

Table 14.2 Some common neurotransmitters and their effects

Neurotransmitter	*Common Effects*
Serotonin (5-HT)	Plays a role in regulating sleep, eating behavior and mood
Dopamine	Involved in motor control, emotional behavior, effects feelings of pleasure and euphoria
Norepinephrine	Regulates mood, arousal, emotions
Acetylcholine	Controls muscular movement
GABA	Primary inhibitory neurotransmitter
Glutamate	Primary excitatory neurotransmitter

closer to threshold, creating a depolarization in the cell. Contrastingly, inhibitory postsynaptic potentials (IPSPs) occur when Cl– enters the cell or K+ leaves the cell. IPSPs move the cell farther away from threshold, or hyperpolarize the cell (Kandel & Siegelbaum, 1991).

Many graded potentials, EPSPs and IPSPs, can influence a cell simultaneously. These potentials sum together to create the potential of the cell at that moment. Postsynaptic potentials can sum over space and time. Spatial summation occurs when two or more postsynaptic potentials affect the cell at two different receptor sites at the same time. Temporal summation occurs when one receptor site is repeatedly stimulated by the same presynaptic cell. Because the threshold of the cell is measured at the axon hillock, the location of the postsynaptic potential is critical. An EPSP near the axon hillock will have a greater effect on driving the cell toward threshold than an EPSP that is far from the axon hillock. Contrastingly, an IPSP close to the axon hillock will have a similar effect by preventing the cell from reaching threshold.

ACTION POTENTIAL

Once threshold has been reached at the axon hillock, voltage-gated Na+ channels open along the initial part of the axon, changing the permeability of the cell to Na+ ions (Figure 14.5). The forces of diffusion and electrostatic pressure drive Na+ into the cell. Remember that diffusion is a factor affected by the concentration gradient. The extracellular fluid has a very high concentration of Na+ ions, therefore Na+ is pushed into the cell. Additionally, Na+ has a positive charge and is attracted to the negative environment of the intracellular fluid. With these two forces working together to push Na+ into the cell, when the voltage-gated channels open Na+ is rapidly forced into the cell. The Na+ channels are open for less than a millisecond; however, enough Na+ ions enter the cell to shoot the potential to +55 mV. This increase in potential causes voltage-gated K+ channels to open. As with Na+, diffusion and electrostatic pressure affect the flow of K+ through the membrane; however, K+ is forced out of the cell. The concentration of K+ is higher inside of the cell, causing K+ to exit the cell. Also, the influx of Na+ into the cell has changed the membrane potential to approximately +55 mV. K+ is positive, therefore the electrostatic pressure pushes K+ out of the cell. As K+ exits the cell, the membrane potential is lowered, resulting in the closing of K+ voltage-gated channels. Because the K+ channels close after the Na+ channels close, the efflux of K+ ions continues even though the membrane potential has been reached. This creates a brief period of hyperpolarization of the membrane. When these K+ ions diffuse into the extracellular fluid, the resting membrane potential is regained (Kuffler & Nichols, 1976).

The propagation of the action potential down the axon is mediated by the myelination of the axon (Katz, 1966). Under the segments of myelin, the axon is not permeable and no ions can pass through the membrane. As shown in Figure 14.6, the Na+ ions diffuse under the myelin segment. Because the ions are more spread out over a larger space, the potential is lowered. Although the charge on the inside of the cell has dropped, it is still high enough to trigger voltage-gated channels to open at the next node. Na+ channels are opened, and sodium rushes into the cell, creating the voltage spike. This influx of Na+ occurs at each node of Ranvier down the length of the axon as the action potential is rejuvenated at each node. This process is called *saltatory conduction,* as the charge looks as if it is jumping down the axon (*saltare* means “leap” or “jump” in Latin).

Once an action potential has been initiated, it will continue until completion. The number of action potentials, or rate of firing of the cell, is partly dependent on the characteristics of the Na+ channels. The absolute refractory period occurs when the Na+ channels have been reset and

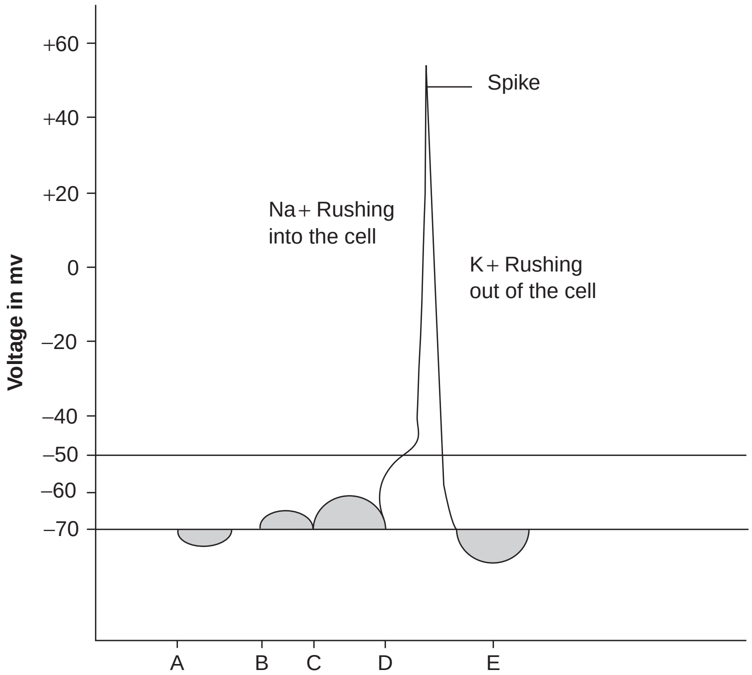

Figure 14.5 Action potential. A: IPSP, bringing the cell farther away from threshold. B and C: EPSPs, summing to bring the cell to threshold (D). E: Brief hyperpolarization following action potential.

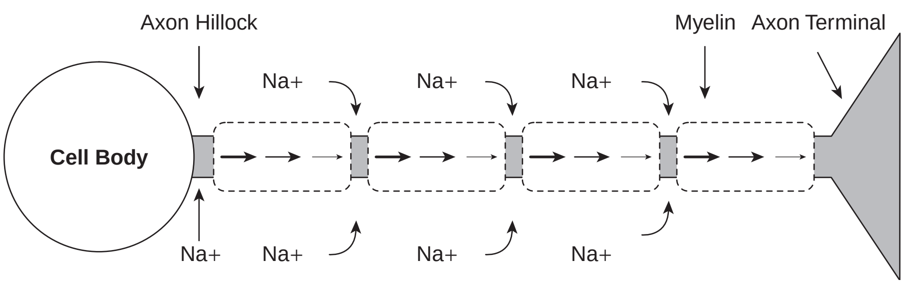

Figure 14.6 Decremental conduction.

they cannot be reopened for a brief (1 ms) period of time (Kalat, 2001). During the absolute refractory period, no stimulus, however great the strength, can initiate an action potential. The relative refractory period occurs because the K+ channels stay open briefly after the Na+ channels have closed, creating a 2–4 ms period of hyperpolarization as K+ still exits the cell (Kalat, 2001). At this point the inside of the cell is more negative than it is at rest (below –70 mV). A stimulus can initiate an action potential during the relative refractory period, but it must be strong enough to drive the cell to threshold.

NEUROTRANSMITTER RELEASE

As the action potential reaches the axon terminals, the increase in potential initiates the opening of voltage-gated channels and calcium ($Ca^{2}+$) enters the cell. The influx of $Ca^{2}+$ causes the synaptic vesicles to migrate to the periphery of the axon terminals and bind to the membrane. The vesicle bound to the terminal membrane forms an omega-shaped complex, and the membrane opens as the neurotransmitter spills into the synaptic gap. This process is called *exocytosis*.

Neurotransmitters excised into the gap migrate across the gap, binding to the receptor sites on the postsynaptic cell. After binding, the neurotransmitter separates from the receptor and returns to the fluid of the cleft. Neurotransmitters can be removed from the synaptic gap in a number of ways. The fate of the neurotransmitter in the gap depends on the neurotransmitter itself. For example, 5-hydroxitryptamine (5-HT), better known as serotonin, is taken back up into the cell and repackaged into synaptic vesicles to be released again. In contrast, acetylcholine is broken down by degrading enzymes into choline and acetate. The choline is taken back up into the cell by specialized protein channels and the acetate is left to disseminate.

SUMMARY

Many different types of neurons all have the same purpose of communicating information from one location (e.g., neuron, muscle, organ) to another. This exchange of information is a complex process called *neurotransmission*. Neurotransmission involves the passing of ions across cell membranes. These ions change the membrane potential, creating EPSPs and IPSPs and bringing the cell closer or farther from an action potential (AP). Once an AP is triggered, a chain reaction occurs, with Na+ entering the cell and K+ leaving the cell. The AP travels down the axon of the neuron signaling the release of neurotransmitters (NT). As the NTs are released into the extracellular space, they open channels on the receiving membrane, letting ions move into or out of the cell and changing the membrane potential.

REFERENCES AND FURTHER READINGS

Bean, R. C., Shepard, W. C., & Chan, H. (1968). Permeability of lipid bilayer membranes to organic solutes. *Journal of Physiology*, *52*, 495–508.

Hille, B. (1967). The selective inhibition of delayed potassium currents in nerve by tetraethylammonium ions. *Journal of Physiology*, *50*, 1287–1302.

Hille, B. (1971). The hydration of sodium ions crossing the nerve membrane. *Proceedings of the National Academy of Sciences, 68,* 280–282.

Hodgkin, A. L., & Huxley, A. F. (1945). Resting and action potentials in single nerve fibres. *Journal of Physiology*, *104*, 176–195.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application of conduction and excitation in nerve. *Journal of Physiology*, *117*, 500–544.

Hodgkin, A. L., & Katz, B. (1949). The effect of sodium ions on the electrical activity in the giant axon of the squid. *Journal of Physiology*, *108*, 37–77.

Julien, R. M. (2001). *A primer of drug action* (9th ed.). New York: Worth Publishers.

Kalat, J. W. (2001). *Biological psychology* (7th ed.). Belmont, CA: Wadsworth/Thompson Learning.

Kandel, E. R. (1991). Nerve cells and behavior. In E. R. Kandel, J. H. Schwartz, & T. M. Jessel (Eds.), *Principles of neural science* (pp. 18–32). Norwalk, CT: Appleton & Lange.

Kandel, E. R., & Schwartz, J. H. (1991). Synaptic transmission mediated by second messengers. In E. R. Kandel, J. H. Schwartz, & T. M. Jessel (Eds.), *Principles of neural science* (pp. 173–193). Norwalk, CT: Appleton & Lange.

Kandel, E. R., & Siegelbaum, S. A. (1991). Directly gated transmission at the nerve-muscle synapses. In E. R. Kandel, J. H. Schwartz, & T. M. Jessel (Eds.), *Principles of neural science* (pp. 135–152). Norwalk, CT: Appleton & Lange.

Katz, B. (1966). *Nerve, muscle, and synapse.* New York: McGraw-Hill.

Koester, J. (1991). *Membrane potential.* In E. R. Kandel, J. H. Schwartz, & T. M. Jessel (Eds.), *Principles of neural science* (pp. 81–94). Norwalk, CT: Appleton & Lange.

Kozorovitskiy, Y., Gross, C. G., Kopil, C., Battaglia, L., McBreen, M., Stranahan, A. M., et al. (2005). Experience induces structural and biochemical changes in the adult primate brain. *Proceedings of the National Academy of Sciences*, *102,* 17479–17482.

Rakowski, R. F., Gadsby, D. C., & De Weer, P. (1989). Stoichiometry and voltage dependence of the sodium pump in voltage clamped, internally dialyzed squid giant axon. *Journal of General Physiology, 93,* 904–941.

Sagar, A., & Rakowski, R. F. (1994). Access channel model for the voltage dependence of the forward-running Na+/K+ pump. *Journal of General Physiology, 104,* 869–894.

Schmitt, F. O., & Schmitt, O. H. (1940). Partial excitation and variable conduction in the squid giant axon. *Journal of Physiology, 98,* 26–46.

Thomas, C. L. (Ed.). (1989). *Taber's cyclopedic medical dictionary* (16th ed.). Philadelphia: F. A. Davis Company

Thompson, R. F. (1993). *The brain: A neuroscience primer* (2nd ed.). New York: W. H. Freeman and Company.

15

Traditional Neuroscience Research Methods

James Kalat

North Carolina State University

> Brain science is difficult and tricky, for some reason; consequently one should not believe a result (one's own or anyone else's) until it is proven backwards and forwards or fits into a framework so highly evolved and systematic that it couldn't be wrong. —*David Hubel*, personal communication, 1992

It might be an interesting exercise, for someone who has nothing better to do, to compile a list of previously held beliefs about the brain that turned out to be false. Neuroscientists have learned to be cautious. Findings need to be replicated under several conditions with several populations or species, and checked with more than one research method. A conclusion that survives a variety of tests is presumably not an artifact of the limitations of any one method or population.

Today, much of the excitement in research comes from studies using functional magnetic resonance imaging (fMRI) and other brain-scan techniques, as discussed in another chapter of this volume. However, these newer procedures supplement the older techniques without replacing them. This chapter surveys a variety of approaches to understanding brain-behavior relations. It will not deal with purely anatomical, chemical, or physiological brain research unrelated to behavior. It also will not provide a how-to-do-it laboratory manual for any of the techniques. The goal is to consider the research strategies, with emphasis on their potentials and limitations.

Research methods for studying brain-behavior relationships fall into four logical categories:

1. Damage part of the brain or decrease its activity and see what deficits occur in behavior.
2. Stimulate increased activity in some brain area and record what behavior increases, or what experience people report.
3. Record brain activity and examine its correlation with simultaneous behaviors.
4. Compare individuals who show unusual features in either their behavior or their brain anatomy, and seek a possible correlation between behavior and anatomy.

Each of these approaches has a variety of subcategories. Some are suitable for laboratory animals but not humans; some provide good temporal resolution but poor spatial resolution, or vice versa. Different methods answer different questions.

EFFECTS OF BRAIN DAMAGE

The earliest discoveries about the functions of various brain areas came from studies of damage. The first clear demonstration that different parts of the nervous system have different functions came in 1822, when François Magendie found that cutting the dorsal nerves of the spinal cord blocked sensory information and cutting the ventral nerves blocked motor output (Gallistel, 1981). In 1861 Paul Broca reported a link between damage to part of the left frontal cortex and a loss of the ability to speak.

Since then, researchers have made countless reports of behavioral impairments after brain damage from stroke, disease, and other causes. The famous studies of patient H. M. provided the first clue linking the hippocampus to memory (Scoville & Milner, 1957). Patients with prosopagnosia demonstrated the importance of the fusiform gyrus for facial recognition (McCarthy, Puce, Gore, & Allison, 1997; Tarr & Gauthier, 2000). We learned about the functions of the corpus callosum from studies of split-brain patients (Gazzaniga, 1970). The list goes on and on.

Many reports of brain damage prompt the reaction, "Isn't it amazing that brain damage can produce such a specific deficit!" For example, people with damage to the parietal lobe can still see objects, but cannot determine their locations. Those with damage to the middle temporal cortex can see in most regards, except that they lose the ability to detect the speed and direction of a moving object (Zihl, von Cramon, & Mai, 1983). Someone with a loss of input to the somatosensory cortex has no conscious sensation of touch, but still reports a "pleasant" experience after a gentle stroke along the skin, showing the possibility of losing the sensation itself but keeping the emotional reaction to it (Olausson et al., 2002).

However, studies of humans face inherent limitations. One is that most brain structures are bilateral, whereas strokes and other spontaneous injuries almost never produce symmetrical, bilateral lesions. Essentially, never do they produce discrete lesions of some of the tiny nuclei of the hypothalamus or amygdala that researchers find so theoretically interesting. Another limitation is that no two people have the same damage, or the same brain functioning before the damage. Thus, neurologists sometimes report patients with similar brain damage but significantly different behavioral symptoms, and the reasons for the differences are difficult to establish. Indeed, even in the classic case of aphasia after damage to Broca's area, neurologists in the century and a half since Broca have made only limited progress in understanding the brain mechanisms of language. The pattern of language abilities and disabilities varies from one patient to another, as does the pattern of brain damage, and only within broad limits can anyone infer the location of brain damage from a patient's behavior, or predict the behavior from the brain damage.

Lesion Studies With Laboratory Animals

Researchers who turn to laboratory animals gain control of the location and the extent of the damage, as well as the animal's history and environment, at the cost of being unable to study language or other distinctively human behaviors. To study the function of a structure on the surface of the brain, researchers can ablate (remove) it with a knife or by suction. To study a structure in the interior of the brain, one procedure is to create a lesion by means of a stereotaxic instrument, which enables precise placement of electrodes in the brain. Stereotaxic atlases, which are available for the brains of many common laboratory animals, specify the position of each brain area in terms of the dorsal-ventral, anterior-posterior, and left-right coordinates, relative to landmarks on the skull of the animal. For example, the ventromedial hypothalamus of a rat is 0.2 mm anterior to bregma (a point on the skull where four major bones join), 10 mm ventral to the dura mater that covers the brain, and 0.8 mm left and right of the midline (Pellegrino & Cushman, 1967). An investigator anesthetizes the animal, shaves part of the scalp and cuts back a flap of skin, drills a small hole in the skull, inserts an electrode insulated except at the tip, and lowers it to the target as determined from the atlas. The experimenter then causes a lesion by passing an electrical current through the tip of the electrode, using a second electrode on the tail or elsewhere as the reference electrode. The experimenter might apply 2 milliamps of direct current for 15 seconds, or a weaker current for a longer duration. An alternative method is to use a radio frequency instead of electricity. In either case, the procedure kills neurons by overheating them. To make bilateral lesions, the researcher repeats the procedure on the other side of the brain. At the end of the experiment, the brain is removed for histological analysis to determine the actual, as opposed to intended, location of the damage. Slices of the brain are subjected to chemicals that stain cell bodies or axons, so that researchers can see the areas of damage. Animals in a control group go through the same procedure except that the experimenter passes no current through the electrode. Once the mainstay of behavioral neuroscience, stereotaxic surgery has become less common as alternative methods have arisen.

An alternative procedure for producing lesions is to inject certain chemicals. For example, an injection of kainic acid, ibotenic acid, or monosodium glutamate into a brain area overstimulates glutamate synapses, leading to a surge of sodium and other positive ions within neurons near the site of injection, thereby poisoning their mitochondria and killing the neurons. The advantage of this procedure over electrolytic surgery is that it spares axons passing through the area. An investigator interested in the lateral hypothalamus, for example, might want to distinguish between the functions of cells there and the many axons passing by (Stricker, Swerdloff, & Zigmond, 1978). Another approach is to inject, either systemically or locally, a chemical that selectively damages particular kinds of neurons or their synapses. For example, 6-hydroxy-dopamine (6-OH-DA), which is absorbed selectively by neurons that release dopamine as their neurotransmitter, attacks the whole network of dopamine neurons while sparing other nearby cells. Similarly, a chemical known as AF64A attacks synapses that use acetylcholine (Sandberg, Sanberg, & Coyle, 1984). Selective damage helps researchers determine how different neurotransmitters contribute to behavior.

In still another approach, known as gene knockout, researchers use biochemical methods to direct a mutation to a particular gene (Joyner & Guillemot, 1994). This procedure can eliminate a particular type of cell, a neurotransmitter, or a receptor. It is well suited to studying a system that is spread out in the brain, instead of destroying a mixture of cell types that happen to be in the same place. However, eliminating a gene sometimes produces widespread and unknown effects that go beyond eliminating one kind of cell or receptor, as the remaining cells reorganize in complex ways that one might not have predicted.

Temporary Inactivation

For many research questions, one might want to study the effect of inactivating a particular brain area at a particular time. For example, is the amygdala more important during acquisition of a learned response, or during its retrieval? To answer such questions, one might insert a thin cannula into a brain area, again using a stereotaxic atlas, and then deliver a compound such as muscimol. Muscimol, a psychoactive chemical found in certain mushrooms, selectively stimulates receptors sensitive to GABA, the brain's main inhibitory neurotransmitter. As a result, muscimol suppresses activity in nearly all neurons in the affected area. A researcher can also inactivate neurons in a particular brain area by cooling them with a metal probe that is cold at the tip but insulated along the rest of its surface (e.g., Chambers & Wang, 2004). Either muscimol or cooling can temporarily suppress activity in a brain area. However, these procedures do not enable someone to turn a brain area off and on quickly. For an example of a study using temporary inactivation, Richard Thompson and his associates inactivated the red nucleus of rabbits during the training phase of a classical conditioning experiment. During this phase, the rabbits showed no conditioned responses. However, when the red nucleus recovered its activity, the rabbits showed the full response, indicating that the red nucleus was necessary for the response but not for learning. In contrast, when the researchers inactivated the lateral interpositus nucleus of the cerebellum, rabbits showed no responses during training and no retention at a later test. Evidently the lateral interpositus nucleus was necessary for learning and not just for the motor response (Krupa, J. K. Thompson, & R. Thompson, 1993).

Researchers would not subject a human brain to either muscimol or a cold metal probe. However, the technique of transcranial magnetic stimulation is available for temporarily inactivating part of the human brain. An investigator applies an intense magnetic field over a portion of the scalp, temporarily inactivating the neurons nearest the magnet (Walsh & Cowey, 2000). Again, this procedure makes it possible to turn a brain area on and off during different phases of a task. A scientific limitation is that it can be applied only to brain areas on the surface, as opposed to the interior of the brain. An ethical limitation is that repetitive pulses can induce seizures and possibly other health problems (Wasserman, 1998).

Difficulties Interpreting Lesion Studies

The results of a lesion study provide, at best, a start toward understanding the function of some brain area or system. They tell us where something happens, but not how. If you remove a chip from inside your computer and find a loss of sound, you know that this chip contributes in some way to producing sound, but you do not know in what way. Similarly, if you damage a brain area and see that an animal learns less readily, you could imagine many possible explanations. The damaged animal might be impaired in perception, attention, motivation, or other processes that affect learning, or in the learning process itself. If learning is impaired, the problem might relate to storing information, retaining it, retrieving it, or avoiding interference that increases forgetting. In some cases, damaging an area impairs some aspect of behavior only because the lesion interrupted axons passing nearby the area (e.g., Almli, Fisher, & Hill, 1979; Berridge, Venier, & Robinson, 1989).

Ideally, researchers like to demonstrate a double dissociation of function—that is, a demonstration that one lesion impairs behavior A more than behavior B, whereas a second lesion impairs B more than A. If some kind of brain damage impairs explicit memory more than implicit memory, we don't know that it is specifically important for explicit memory, because an alternative explanation is that explicit memory is more difficult than implicit memory. In that case, any lesion that impaired attention, arousal, motivation, or other processes would impair explicit more than implicit memory. However, if other kinds of brain damage impair implicit more than explicit memory, we gain confidence that the two kinds of lesions are really affecting memory in different ways.

A further difficulty with lesion studies is that the brain's plasticity can mask the effect of a deficit. For example, many neurotransmitters contribute to the regulation of feeding, but a targeted disruption of one of them may have little effect because the others compensate. In some cases, disrupting one pathway sets in motion a reorganization of others with far-reaching, unpredicted effects. In one study, researchers knocked out the gene for one type of potassium channel in the neuronal membrane in mice. Although one would expect to find deficits, the first effect they noticed was that the mice had a thousandfold *increase* in their olfactory sensitivity (Fadool et al., 2004).

EFFECTS OF BRAIN STIMULATION

If brain damage impairs some behavior, stimulation should increase it. Thus, one way to confirm the results from a lesion study is to follow up with a stimulation study. One method suitable for use with healthy humans is to apply a magnetic field to the scalp, using much briefer and milder amounts than with the transcranial magnetic stimulation mentioned earlier. The result is increased activity of neurons in the brain areas nearest the magnet (Fitzgerald, T. L. Brown, & Daskalakis, 2002). Occasionally, neurologists have electrically stimulated spots in the brain of a person who is undergoing brain surgery under local anesthesia. For example, that kind of research was responsible for the first mapping of the human motor cortex and somatosensory cortex (Penfield & Rasmussen, 1950).

With laboratory animals, researchers can use a stereotaxic instrument to insert an electrode to stimulate a small area anywhere in the brain, such as the hypothalamus. Alternatively, they inject chemicals either into the bloodstream or directly into the brain. Examples include drugs that stimulate or block particular neurotransmitter receptors. Other examples include hormones, which modulate activity of whatever neurons respond to them. If injected systemically, the chemicals exert their effects wherever the corresponding receptors occur in the brain. If injected in small amounts into a specific brain area, they have much more localized effects. This type of research has become common in behavioral neuroscience.

An alternative to injecting chemicals is to use a virus vector to increase the expression of a particular gene. For example, male prairie voles form pair bonds with females and help them rear their young, whereas males of the closely related meadow voles do not. They differ in their secretion of the pituitary hormone vasopressin, with prairie voles secreting much more. When researchers used a viral vector to transfer a gene into the hypothalamus and pituitary of meadow voles, increasing their production and release of vasopressin, the males became monogamous and began helping with infant care (Lim et al., 2004).

Directly stimulating the brain is, of course, unlike normal brain activity. When you see a face or hear a melody, you have a distinct pattern of activity distributed over a huge number of cells, modulated in a specific way over time. Electrical or magnetic stimulation of a small brain area can produce meaningful results in certain areas where precise timing is less important—for example, the areas controlling hunger or thirst. However, stimulation of any one spot in the primary visual cortex produces reports of flashes of light, not of meaningful objects.

For decades, the same appeared to be true of the motor cortex as well. Brief electrical stimulation of monkeys' motor cortex produced only uncoordinated muscle twitches. However, later researchers found that a half-second pulse of stimulation elicited complex sequences of actions. In one case, stimulation of a certain spot led a monkey to move its hand from wherever it was to its mouth, even though the monkey had to use different muscles depending on the starting position of the hand (Graziano, Taylor, & Moore, 2002). In this case, stimulation research led us to a new conclusion: Cells of the motor cortex have their output organized according to the outcome of the movement, not the muscle contractions themselves (Scott, 2004).

RECORDING BRAIN ACTIVITY

We generally think of biological psychology or behavioral neuroscience as a field that uses experimental methods. However, a significant amount of research also uses correlational methods. For example, researchers correlate observed behavior with measurements of brain activity.

Electroencephalography and Related Procedures

In contrast to research that requires inserting electrodes into the brain, subjecting it to hazardous chemicals, or applying strong magnetic fields, one of the least intrusive research methods is an electroencephalograph (EEG). To set up an EEG, a researcher glues some electrodes to a person's scalp—ranging from just a few electrodes to a hundred or more. Each electrode provides an average for the instantaneous activity of the population of cells under it. That output is amplified and recorded. A classic use of EEG is to monitor stages of sleep, with slow waves indicating stages 3 and 4 of sleep.

Researchers can use the same equipment to record brain activity in response to sensory stimuli, in which case the procedure is called *evoked potentials* or *evoked responses*. Researchers attend particularly to the responses evoked at particular latencies after a stimulus. For example, a typical person shows a strong positive response, known as the P300 wave, about 300 ms after an especially meaningful stimulus, such as a rare target to which the person is supposed to respond. The P300 wave is weaker than average among people with schizophrenia, attention deficit disorder, and other conditions that impair attention (Du et al., 2006; Price et al., 2006).

A magnetoencephalograph (MEG) is similar to an EEG, but it measures the faint magnetic fields generated by brain activity instead of electrical activity (Hari, 1994). MEG has superb resolution over time, sensitive to changes occurring in as little as a millisecond. This fine resolution enables researchers to follow a wave of brain activity from its point of origin to all the other areas that it affects (Salmelin, Hari, Lounasmaa, & Sams, 1994). However, both EEG and MEG have mediocre resolution over space, as they average the activity occurring over a surface of at least a square centimeter.

Body Fluid Samples

Because the brain's metabolites eventually wash out of the brain, researchers can monitor body fluids as an admittedly imperfect but relatively simple gauge of certain kinds of brain activity. For example, when a neuron releases the neurotransmitter serotonin, much of the serotonin returns to the presynaptic neuron, but some of it is metabolized to 5-hydroxyindoleacetic acid (5-HIAA). The neuron synthesizes new serotonin molecules to replace those it loses as 5-HIAA; the overall process is called *serotonin turnover*. Researchers can assay the

5-HIAA levels of the cerebrospinal fluid, blood, or urine to estimate the amount of serotonin turnover in the brain. Although this method is far from perfect, as platelets and other peripheral tissues produce serotonin also, the measurements do correlate with certain aspects of behavior. For example, decreased levels of serotonin turnover correlate with increased probability of violent behaviors in rats, monkeys, and humans (T. L. Brown et al., 1982; Higley et al., 1996; Valzelli, 1973). This approach is limited to the study of fairly long-lasting trends in behavior. Because the half-life of 5-HIAA in the cerebrospinal fluid is more than an hour (Mignot, Laude, & Elghozi, 1984), researchers cannot apply this measure to moment-by-moment changes in behavior.

Histological Measurements

Another way to monitor activity in various brain areas in laboratory animals is to take slices of the brain and apply stains that react to the protein *fos*. The *fos* gene has pervasive effects throughout the body, being activated in the early stages of cell division as well as during neural activity. Within neurons, any pattern of sustained activity, such as might occur during learning, for example, leads to an increase in *fos* expression that researchers can detect beginning about 15 minutes later. It is apparently a reaction to the increased neuronal activity, rather than a prerequisite for starting it. Although the slow time course precludes measurement of rapid changes, the method does reveal changes over a longer time scale. In one study, researchers found that seizures increased *fos* expression in the dentate gyrus of the hippocampus 90 minutes later, and still later in surrounding areas of the hippocampus and cerebral cortex (Morgan, Cohen, Hempstead, & Curran, 1987). Staining for *fos* activity has been useful for such questions as identifying the brain areas responsive to social stress in rodents (Martinez, Calvo-Torrent, & Herbert, 2002) and specifying the nuclei of the hypothalamus that respond to cholecystokinin, a satiety signal (Kobelt et al., 2006).

CORRELATING BRAIN ANATOMY WITH BEHAVIOR

In addition to correlating behavior with brain activity, researchers sometimes correlate behavior with brain anatomy. That approach has persisted throughout the history of neurology, although the earliest examples are not to be celebrated. Franz Gall and the other phrenologists of the 1800s attempted to relate people's personalities to their brain anatomy, using bumps and depressions on the skull as an indicator of the brain structures underneath. This effort was ill-fated from the start, as the surface of the skull is a poor indicator of brain anatomy. Furthermore, the phrenologists used their data uncritically, sometimes drawing conclusions about the functions of a particular brain area from their observations of just one or a few people.

Another early goal was to relate intelligence to brain size or structure. Several societies arose in the 1800s and 1900s devoted to preserving the brains of eminent men after death, including the members of those societies themselves, in hope of finding some distinctive aspect of the size or structure of the brain that separated successful from less successful men. (Women were ignored.) These projects led to no conclusion, partly because achieving eminence depends as much on opportunity as intellectual ability, and partly because the correlation of brain anatomy with intelligence is at best a weak one, not obvious with visual inspection of casually selected samples (Burrell, 2004; Schoenemann, Budinger, Sarich, & Wang, 2000).

Although we now regard these early pursuits with either disdain or amusement, let's admit that we are still subject to a similar temptation today. When the famous scientist Albert Einstein died in 1955, researchers removed his brain, hoping to find some unusual feature responsible for his brilliance. Decades later, the researchers—who were apparently in no hurry about this task—reported that Einstein had a higher than average ratio of glia to neurons in one brain area (Diamond, Scheibel, Murphy, & Harvey, 1985), and that his inferior parietal cortex was noticeably expanded (Witelson, Kigar, & Harvey, 1999). Although the researchers were careful not to draw conclusions, the implication was that these brain features had some relation to Einstein's scientific achievements. Perhaps so, but we cannot be sure until we compare the brains of many other scientists of Einstein's stature—and let's not hold our breath waiting. We also need to determine how often such features occur in the brains of less distinguished people. The problem is not just that Einstein was a single case, but also that the researchers examined many aspects of his brain in search for something noteworthy. Given that we do not know exactly how many total hypotheses researchers considered, we cannot readily interpret a couple of differences between Einstein's brain and everyone else's.

Nevertheless, much useful research does use the strategy of correlating brain anatomy with behavior. Fascinating studies have reported expansion of the right auditory cortex in professional musicians (Schneider et al., 2002), increased representation of the fingers of the left hand in the somatosensory cortex of lifelong violin players (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995), and expansion of the hippocampus in London taxi drivers (Maguire et al., 2000). Species comparisons have found that members of the jay family with the greatest spatial memory also have the largest hippocampus (Basil, Kamil, Balda, & Fite, 1996). Each of these differences is consistent enough across individuals to be reliable, and each confirms what other kinds of research indicate about the functions of those brain areas.

If we want to study the physiological basis of an uncommon psychological disorder, the only reasonable starting point is to compare the brains of people affected with the disorder to those of other people. For example, Williams syndrome is a genetically determined condition marked by mental retardation in most regards, yet often remarkably good skills at language, music, and social relationships. The syndrome is consistently associated with reduced gray matter in the occipital cortex and parts of the thalamus, but greater than average volume in parts of the temporal and frontal cortices (Reiss et al., 2004). These differences are reasonably consistent over a large number of people, and they have the potential to inspire further research. For present purposes, the point is that the research is necessarily correlational in nature.

Researchers who study schizophrenia, depression, and a host of other conditions compare the brains of individuals with the condition to control participants. In each case, moderately consistent patterns do emerge, which may provide a clue to understanding the disorder. The most important way in which this approach differs from phrenology is that modern researchers compare large groups of people, reporting only the differences that emerge as statistically reliable. A lingering problem is that anyone who examines all parts of the brain is simultaneously testing a huge number of hypotheses, and therefore has a strong possibility of finding differences by accident.

An additional problem with this kind of research is the choice of a control group. Suppose, for example, we are searching for distinctive brain features associated with schizophrenia. We assemble a group of people with schizophrenia. Who should be in the control group? Any good researcher knows to match people for age and sex, but is that enough? In many cases the ideal comparison is between siblings, of whom one has the condition and one does not. However, in the case of schizophrenia, none of these approaches is fully satisfactory. Many people, especially men with schizophrenia, abuse alcohol or other drugs. Given that alcohol and other drugs can affect brain anatomy, some of the brain abnormalities associated with schizophrenia may in fact be consequences of the drugs themselves, or of drug-related problems, including poor diet and head trauma (Sullivan et al., 2000). If we want to examine the effects of schizophrenia itself, we need a control group matched for use of alcohol and other drugs.

BEHAVIORAL STUDIES

Finally, consider one additional research approach, which does not tell us anything new about brain-behavior relations, but uses what we do know to provide information about individuals. If research has already established that a certain kind of brain damage leads to a particular behavioral deficit, researchers can test that kind of behavior to draw inferences about possible brain impairments. For example, in the Wisconsin General Test Apparatus, one task is to sort a set of cards first by one rule (such as color) and then by a second rule (such as number). Any damage or impairment of the prefrontal cortex impedes a person's ability to shift to the second rule. Many people with schizophrenia perform poorly on this task, implying possible difficulties in their prefrontal cortex.

SUMMARY

The Nobel laureate biologist Sydney Brenner was quoted as saying that progress in science depends on "new techniques, new discoveries, and new ideas, probably in that order" (McElheny, 2004, p. 71). That is, most scientific discoveries begin with new or improved methods of measuring something. Brain science relies on a huge variety of methods, including the ones described here, which differ enormously in their details but generally fall into four categories: the effects of brain damage, the effects of brain stimulation, measurements of brain activity during behavior, and correlations of brain structure with behavior. Future research will advance as we devise and improve new methods of measurement.

REFERENCES AND FURTHER READINGS

Almli, C. R., Fisher, R. S., & Hill, D. L. (1979). Lateral hypothalamus destruction in infant rats produces consummatory deficits without sensory neglect or attenuated arousal. *Experimental Neurology, 66,* 146–157.

Basil, J. A., Kamil, A. C., Balda, R. P., & Fite, K. V. (1996). Differences in hippocampal volume among food storing corvids. *Brain, Behavior and Evolution, 47,* 156–164.

Berridge, K. C., Venier, I. L., & Robinson, T. E. (1989). Taste reactivity analysis of 6-hydroxydopamine-induced aphagia: Implications for arousal and anhedonia hypotheses of dopamine function. *Behavioral Neuroscience, 103,* 36–45.

Brown, G. L., Ebert, M. H., Goyer, P. F., Jimerson, D. C., Klein, W. J., Bunney, W. E., et al. (1982). Aggression, suicide, and serotonin: Relationships of CSF amine metabolites. *American Journal of Psychiatry, 139,* 741–746.

Burrell, B. (2004). *Postcards from the brain museum*. New York: Broadway Books.

Chambers, K. C., & Wang, Y. (2004). Role of the lateral parabrachial nucleus in apomorphine-induced conditioned consumption reduction: Cooling lesions and relationship of c-Fos-like immunoreactivity to strength of conditioning. *Behavioral Neuroscience, 118,* 199–213.

Diamond, M. C., Scheibel, A. B., Murphy, G. M., & Harvey, T. (1985). On the brain of a scientist: Albert Einstein. *Experimental Neurology, 88,* 198–204.

Du, J., Li, J., Wang, Y., Jiang, Q., Livesley, W. J., Kerry, L. J., et al. (2006). Event-related potentials in adolescents with combined ADHD and CD disorder: A single stimulus paradigm. *Brain and Cognition, 60,* 70–75.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science, 270,* 305–307.

Fadool, D. A., Tucker, K., Perkins, R., Fasciani, G., Thompson, R. N., Parsons, A. D., et al. (2004). Kv1.3 channel gene-targeted deletion produces "super-smeller mice" with altered glomeruli, interacting scaffolding proteins, and biophysics. *Neuron, 41,* 389–404.

Fitzgerald, P. B., Brown, T. L., & Daskalakis, Z. J. (2002). The application of transcranial magnetic stimulation in psychiatry and neurosciences research. *Acta Psychiatrica Scandinavica, 105,* 324–340.

Gallistel, C. R. (1981). Bell, Magendie, and the proposals to restrict the use of animals in neurobehavioral research. *American Psychologist, 36,* 357–360.

Gazzaniga, M. S. (1970). *The bisected brain.* New York: Appleton-Century-Crofts.

Graziano, M. S. A., Taylor, C. S. R., & Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron, 34,* 841–851.

Hari, R. (1994). Human cortical functions revealed by magnetoencephalography. *Progress in Brain Research, 100,* 163–168.

Higley, J. D., Mehlman, P. T., Higley, S. B., Fernald, B., Vickers, J., Lindell, S. G., et al. (1996). Excessive mortality in young free-ranging male nonhuman primates with low cerebrospinal fluid 5-hydroxyindoleacetic acid concentrations. *Archives of General Psychiatry, 53,* 537–543.

Joyner, A. L., & Guillemot, F. (1994). Gene targeting and development of the nervous system. *Current Opinion in Neurobiology, 4*, 37–42.

Kobelt, P., Paulitsch, S., Goebel, M., Stengel, A., Schmidtmann, M., van der Voort, I. R., et al. (2006). Peripheral injection of CCK-8S induces Fos expression in the dorsomedial hypothalamic nucleus in rats. *Brain Research, 1117,* 109–117.

Krupa, D. J., Thompson, J. K., & Thompson, R. F. (1993). Localization of a memory trace in the mammalian brain. *Science, 260,* 989–991.

Lim, M. M., Wang, Z., Olazábal, D. E., Ren, X., Terwilliger, E. F., & Young, L. J. (2004). Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. *Nature, 429,* 754–757.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences, 97,* 4398–4403.

Martinez, M., Calvo-Torrent, A., & Herbert, J. (2002). Mapping brain response to social stress in rodents with *c-fos* expression: A review. *Stress, 5,* 3–13.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9,* 605–610.

McElheny, V. (2004). Three Nobelists ask: Are we ready for the next frontier? *Cerebrum, 5,* 69–81.

Mignot, E., Laude, D., & Elghozi, J. L. (1984). Kinetics of drug-induced changes in dopamine and serotonin metabolite concentrations in the CSF of the rat. *Journal of Neurochemistry, 42,* 819–825.

Morgan, J. I., Cohen, D. R., Hempstead, J. L., & Curran, T. (1987). Mapping patterns of c-*fos* expression in the central nervous system after seizure. *Science, 237,* 192–197.

Olausson, H., Lamarre, Y., Backlund, H., Morin, C., Wallin, B. G., Starck, G., et al. (2002). Unmyelinated tactile afferents signal touch and project to insular cortex. *Nature Neuroscience, 5,* 900–904.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man.* New York: Macmillan.

Pellegrino, L. J., & Cushman, A. J. (1967). *A stereotaxic atlas of the rat brain*. New York: Appleton-Century-Crofts.

Price, G. W., Michie, P. T., Johnston, J., Innes-Brown, H., Kent, A., Clissa, P., et al. (2006). A multivariate electrophysiological endophenotype, from a unitary cohort, shows greater research utility than any single feature in the Western Australian family study of schizophrenia. *Biological Psychiatry, 60,* 1–10.

Reiss, A. L., Eckert, M. A., Rose, F. E., Karchemskiy, A., Kesler, S., Chang, M., et al. (2004). An experiment of nature: Brain anatomy parallels cognition and behavior in Williams syndrome. *Journal of Neuroscience, 24,* 5009–5015.

Salmelin, R., Hari, R., Lounasmaa, O. V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature, 368,* 463–465.

Sandberg, K., Sanberg, P. R., & Coyle, J. T. (1984). Effects of intrastriatal injections of the cholinergic neurotoxin AF64A on spontaneous nocturnal locomotor behavior in the rat. *Brain Research, 299,* 339–343.

Schneider, P., Scherg, M., Dosch, G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5,* 688–694.

Schoenemann, P. T., Budinger, T. F., Sarich, V. M., & Wang, W. S.-Y. (2000). Brain size does not predict general cognitive ability within families. *Proceedings of the National Academy of Sciences, 97,* 4932–4937.

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience, 5,* 532–544.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20,* 11–21.

Stricker, E. M., Swerdloff, A. F., & Zigmond, M. J. (1978). Intrahypothalamic injections of kainic acid produce feeding and drinking deficits in rats. *Brain Research, 158,* 470–473.

Sullivan, E. V., Deshmukh, A., Desmond, J. E., Mathalon, D. H., Rosenbloom, M. J., Lim, K. O., et al. (2000). Contribution of alcohol abuse to cerebellar volume deficits in men with schizophrenia. *Archives of General Psychiatry, 57,* 894–902.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by experience. *Nature Neuroscience, 3,* 764–769.

Valzelli, L. (1973). The "isolation syndrome" in mice. *Psychopharmacologia, 31*, 305–320.

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience, 1*, 73–79.

Wasserman, E. M. (1998). *Risk and safety of repetitive magnetic stimulation.* Report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation, June 5–7, 1996.

Zihl, J., von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain, 106,* 313–340.

16

IMAGING TECHNIQUES FOR THE LOCALIZATION OF BRAIN FUNCTION

BRENDA ANDERSON
Stony Brook University

The incorporation of human imaging methods into the psychological sciences has been rapid and dramatic. The movement of behavioral research in this direction requires careful consideration of these methods and what they reveal about behavior and the brain. For this reason, the present chapter will highlight the basic assumptions that underlie these methodologies. The assumption that behavior is localized in the brain underlies the need for imaging methods in psychology. But these methods are only as good as the information they provide. Accordingly, we start with a review of the relation between neural activity and metabolism in the brain, and then proceed to describe the relation between blood flow and the signals measured, first with functional magnetic resonance imaging (fMRI). Because both positron emission tomography (PET) and fMRI utilize structural magnetic resonance (MR) images to identify the location of the activation, I will briefly discuss the basis of structural MRI and will illustrate its use to psychologists. The majority of the chapter will describe the basis of fMRI, which is the imaging method most widely used at the moment by psychologists. For comparison purposes, there will be a brief description of PET functional imaging. Finally, I discuss the limits of these methodologies in relation to the questions of interest in psychology.

ELEMENTAL FUNCTIONS LOCALIZED WITHIN THE BRAIN

Behind the emergence of human imaging methods is the belief that function is localized in the brain. The belief that the mind and behavior are products of a biological machine, the brain, has emerged slowly over the last several centuries. These beliefs began with Galileo and Newton's interpretation that the world is mechanical in nature (Schultz, 1981). Once Newton established that the laws of nature extended to the stars and planets, philosophers began to question why the body and mind should be excluded from these laws. Ultimately, the empiricists began to believe that the mind is a product of the natural world, and therefore can be studied like the natural world. The empiricists concluded that thoughts are based solely on sensory experiences and associations between sensory experiences. Eventually they reasoned that sensory impressions and associations are formed within the brain. Altogether, this progression of philosophical ideas forms the early basis of our current belief that the mind and its output, behavior, are a product of the brain.

In the field of neuroscience, there was an initial conflict as to whether specific locations within the brain subserved specific behaviors, or whether the brain operated as a whole (the aggregate field view of Fluorens; Kandel, Schwartz, & Jessell, 2000). Over time, there developed a set of data that together pointed squarely to the concept that elemental functions, a reduction from the personality traits proposed to be localized by Gall, are localized within discrete regions of the brain. Broca and Wernicke, for example, found that deficits in language were associated with damage to specific brain locations. Fritsch and Hitzig were able to stimulate specific motor responses by electrically stimulating the motor cortex. Brodmann demonstrated that despite the six layers of the cerebral cortex, there are subtle differences in the cellular and layer architecture (cytoarchitectonics) over the cortex. Thus the potential for differential computational processing across cortical regions existed. Eventually, the field of neuroscience has adopted and framed data on the belief that function is localized within the brain.

Prior to the advent of human imaging, the field of neuropsychology utilized natural lesions to develop an understanding of the relation between brain lesions and behavioral deficits. Only rarely were there other opportunities to obtain converging data by measuring neural activity. Typically, electrophysiological recordings are carried out only in epileptics preparing for surgery, and these studies are problematic because the population may not necessarily represent the population as a whole. Furthermore, only a few brain sites can be recorded from at any one time. To avoid invasive methods, researchers have used recordings of scalp EEG widely. Although these methods, like electrophysiological recordings from implanted electrodes, have a superior ability to measure potentials over very brief time periods (ms), the EEG potentials could not be localized. Instead they reflect electrical activity summed over large areas. Given that function is localized, the location of the source of activity is of great interest. Human imaging methods brought about the opportunity to study location of function without the limitations of natural dysfunction or disease. Most of these methods, however, reflect signals that are collected over seconds. Accordingly, there is strong interest in combining the location information from human imaging methods and rapid electrical activity provided by EEG recordings.

The advent of human imaging methods, PET in the 1980s and fMRI in the early 1990s, provided greater opportunities to visualize activation in areas of the brain during sensory, perceptual, cognitive, social, affective, and motor processing. Furthermore, fMRI allowed the study of activation throughout the whole brain rather than in single regions at one time. The potential to image many areas at once opens up the possibility of studying how areas work together to produce complex behaviors.

COUPLING BETWEEN NEURAL ACTIVITY, METABOLISM, AND BLOOD FLOW

Although PET and fMRI are two different methods for studying human neural activation during task performance, these techniques rely on the assumption that increases in brain metabolism are reflections of an increase in neural activation. Therefore, to trust the relevance of the data from these methodologies requires an understanding of the relation between neural activity and brain metabolism. It may help to know first that the brain uses a substantial proportion of the energy consumed by the body. The brain uses 15 percent to 20 percent of the oxygen consumed despite making up only 2 percent of body weight (Clarke & Sokoloff, 1999; Roland, 1993). The glucose and oxygen used by the brain serves to produce ATP. This molecule is valuable because the free energy it stores is easily accessible. ATP is considered the currency required for the biological reactions that support life. Instead of storing ATP, the brain makes ATP as needed. This requires that neural activity stimulate increases in metabolism. In other words, as neural activity increases, it must produce signals to accelerate multiple stages of metabolism necessary to produce ATP. For example, as neuronal activity increases, it produces signals that increase local blood flow (Roland, 1993). These signals cause dilation of upstream arterioles, leading to an increase in blood, and the oxygen and glucose it carries, in local capillary beds. Capillaries, the smallest vessels, transport oxygen and glucose into brain tissue. Glucose is used for glycolysis, which produces pyruvate, lactate, and ATP. The ATP produced from glycolysis doesn't adequately meet the needs of neural activity. The pyruvate and lactate, however, are used for a second stage of metabolism, the Kreb's (TCA) cycle. The Kreb's cycle, in turn, produces high-energy compounds needed for the final stage of metabolism, oxidative phosphorylation. This last stage requires oxygen and produces the greater amounts of ATP needed to support brain activity. In summary, when neural activity increases, it produces signals that increase the production of ATP. The ATP is broken down to release the free energy needed to restore the resting neuronal state.

It has long been recognized that increased neuronal activity can increase blood flow (Roy & Sherrington, 1890), but the biological signals responsible for this "neurovascular coupling" remain a mystery (Metea & Newman, 2006; Roland, 1993). Over the years, many signals have been suspected. These include CO_2, pH, extracellular potassium, purines (adenosine, AMP, ADP, and ATP), and more recently, EETs and 20-HETE, prostoglandins, and cyclooxygenase metabolites (Metea & Newman, 2006; Roland, 1993). Each has substantial support for a role in mediating neurovascular coupling, but also fails in some experimental models to cause vasodilation. The ongoing contradictions in experimental outcomes likely reflect different mechanisms operating in different regions of the brain, thus leading to different outcomes in different experimental models, and also the likelihood that several mechanisms work in concert to alter blood flow, making the effectiveness of any single signal difficult to detect. For the present purpose, it is simply important to note that within the neuroscientific community it is widely accepted that products of neuronal activity (which are yet to be agreed upon) cause vasodilation. Other experimental data support a coupling between neural activity and glucose uptake (Sokoloff et al., 1977), glycolysis (Roland, 1993), oxygen consumption (Hill & Howarth, 1958), and oxidative phosphorylation (Wong-Riley, 1989). Because of this tight coupling, the activity of multiple stages of metabolism can be used as indirect indicators of neuronal activity. This coupling is exploited by many imaging methods. fMRI measures relative changes in blood flow and volume, whereas PET imaging is often used to quantify changes in glucose and oxygen uptake. In summary, although PET and fMRI provide *indirect* measures of neural activity, the basis for the relation between metabolism and neural activity is well established. Furthermore, metabolic mapping based on these relationships was used in neuroscience well before the advent of human imaging methods.

CORRELATION OF SYNAPTIC POTENTIALS WITH METABOLISM AND BLOOD FLOW

At the level of a single neuron, it is of interest to consider whether the metabolic changes being measured reflect synaptic input or action potential output. Neurons are made up of a number of compartments. The compartments include the dendrites, which serve to greatly expand the membrane available for synaptic input, and the cell body, which both serves as a site for synaptic input and stores genetic material. The axon hillock arising from the cell body integrates the synaptic input, and when that input reaches a threshold, it produces an action potential. Action potentials are conducted by the axon over long distances at a rapid rate and without amplitude loss. Action potentials arrive at the synaptic terminals, which are responsible for releasing transmitter that serves to relay chemical information to the postsynaptic cells. This chemical information causes an electrical change in the postsynaptic membrane (dendrite) in the form of a “synaptic potential.” Synaptic potentials are difficult to detect with electrical recording methods. Direct measures of neural activity (i.e., electrophysiological recordings) often represent action potentials, because the potential amplitude of action potentials is far larger than that of synaptic potentials, and therefore easier to detect. Action potentials are far shorter in duration (1–10 ms) than synaptic potentials (5 ms to 20 minutes; Kandel et al., 2000). The majority of ATP serves to maintain membrane potentials. Thus, it is relevant to ask, does the increase in metabolism, which produces ATP, support the production and conduction of action potentials, or the enormous number of synaptic potentials on the vast surface of the dendritic membrane? Said more simply, does metabolism correlate more closely with synaptic input or action potentials output?

Investigation of the location of highest metabolic demand suggests that synaptic potentials create more metabolic demand than action potentials. Capillary density increases as synaptic density increases (Dunning & Wolff, 1937; as cited by Roland, 1993), and enzymes coupled to the production of ATP, along with the organelles that house them, are also highest in dendritic fields (Borowsky & Collins, 1989; Erecinska & Wilson, 1982; Zheng, LaMantia, & Purves, 1991). The highest energy consumption occurs in regions with the highest density of synapses, not regions with cell bodies, where action potentials would originate. For example, stimulation of rat whiskers increased the metabolism of glucose in areas of the cortex that represented whiskers. The cortical layers with the highest increase in metabolism were those layers with the highest density of dendrites and synapses (Hand, Greenberg, Sylvestro, Weller, & Reivich, 1979). Similar results were found from the stimulation of monkey fingertips (Juliano, Hand, & Whitsel, 1981). Taking advantage of the gross segregation of cell bodies in the dorsal root ganglion near the spinal cord, Kadekaro, Crane, and Sokoloff (1985) first showed that baseline levels of glucose utilization differed between areas containing dendrites and synapses compared to areas containing cell bodies only. Further, as stimulation was increased over a range of 0 to 15 Hz, glucose utilization increased only in the regions with synapses, not in the region with cell bodies (Kadekaro et al., 1985). In conclusion, neurovascular coupling appears to closely represent metabolism influenced by synaptic input rather than by action potential production and conduction. Thus, to integrate single unit recordings from nonhuman primates and human imaging data requires an understanding that single units would be expected to correspond to metabolic change in the downstream dendritic fields targeted by axonal projections.

With an understanding of the physiological bases supporting the use of metabolic measures as indicators of neural activity, it is possible to proceed to a description of imaging methods that are currently applied in human subjects. Both PET and MR imaging of functional activation utilize the superior spatial information provided by MR structural images. For that reason, we will start with a brief description of the basis of MRI, and its use for the acquisition of images of brain structure.

ORIGIN AND BASIS OF MRI

Prior to *imaging* with magnetic resonance, researchers used nuclear magnetic resonance (NMR) to study the structure of molecules. In a magnetic field, atoms will absorb radio frequencies (RF) and resonate. Each atom has its own resonant frequency that will be influenced by the magnetic field. The resonating frequency then codes which atoms are present, and in what number, in a molecule. The integration of knowledge from each of the resonant frequencies can lead to an understanding of the molecular structure. Such studies typically use pure molecules in solution, and are important in the field of organic chemistry, but less so for the field of biology.

NMR exploded into biology and medicine when methods were developed that allowed the location of the resonant frequencies to be mapped in space. This method is widely known as *magnetic resonance imaging,* or MRI. Because the resonating frequency of an atom is dependent in part on the strength of the magnetic field, a variation in field strength over space can be used to influence the resonating frequencies (Schild, 1990). The frequencies can, in turn, be used to decode the position in space from which they arise. By applying magnetic gradients in three different dimensions, it is possible to identify the origin of a signal in three-dimensional space. Suddenly, NMR was no longer limited to the collection of spectral frequencies from homogenous solutions, but instead could be used to collect three-dimensional images of atoms in space. MRI has been widely used to noninvasively image the human body.

MRI works by taking advantage of signals produced by spinning protons (Pykett, 1982). Because protons have an electrical charge that is moving, they produce an electrical current that can be detected. Because electrical currents produce a magnetic field, each proton is equivalent to a small magnetic bar. In MRI, a longitudinal magnetic field is applied. The mobile protons orient the axis of their spin with that of the longitudinal magnetization, like spinning tops aligned by gravity. Rather than spin perfectly parallel to this axis, they wobble around it, much like a wobbling top. Two variables are of interest for MRI: first, the vertical orientation of the tops (which side is up) and their low angle of spin around the axis, and second, whether they wobble together on one side of the axis, or randomly around the axis (Pykett, 1982; Schild, 1990). The protons produce an RF signal that a receiver coil in the MRI can detect. For the psychologist, the three variables used for imaging the human brain include T_1, T_2 and T_2^*. To understand what is measured by the MRI, see Figure 16.1. Briefly, the spinning protons wobble randomly around a central axis. There is only a small angle between their orientation and the longitudinal axis. When a brief RF pulse is applied, their angle relative to the axis increases (they tip by 90 or 180 degrees). Furthermore, the RF pulse causes the protons to spin in synchrony from one side of the axis to another (called *coherence of precession*). Over time, coherence is lost, which is called *dephasing*. The dephasing reflects inhomogeneities in the magnetic field produced by both internal sources (e.g., tissue) and external sources (e.g., iron in blood). The time it takes to dephase is measured as T_2^* and T_2. The time it takes for the protons to return to their low angle of spin around the longitudinal axis is referred to as T_1. T_1 is short for lipids (dense in white matter) and long for liquids (e.g., cerebral spinal fluid). Accordingly, the signal mapped across an image of the brain would be light for white matter, dark for gray matter (relatively more water content), and darkest for CSF. MRIs designed to image brain structure are usually weighted to measure either T_1 or T_2. Liquids have long T_2 relaxation times (opposite T_1). The T_2 weighted image of the brain would have dark white matter with its dense lipids, lighter gray matter with relative less density of lipids and greater water content, and lightest for the cerebral ventricles.

USES OF STRUCTURAL MRI

The ability of T_1 and T_2 weighted images to distinguish between gray and white matter, based on differences in water content, offers the opportunity to study differences in relative brain region size. Size can be studied across individuals, or within individuals over time. The measurement of anatomical differences is of particular interest for understanding disorders such as depression (Videbech & Ravnkilde, 2004; Vythilingam et al., 2004), schizophrenia (Davatzikos et al., 2005; DeLisi, Sakuma, Maurizio, Relja, & Hoff, 2004; Schild, 1990), and PTSD (Pitman et al., 2006). Differences have also been studied, in normal humans with exceptional talents such as musical ability (Gaser & Schlaug, 2003). Because structural MRIs can be taken repeatedly, they are used to study whether experience can lead to changes in brain region size over time. One such study demonstrated that the brain area sensitive to visual motion (areaMT) increased in size following the acquisition of a skill for juggling (Draganski et al., 2004; Draganski et al., 2006), which presumably requires visual as well as motor learning.

Structural MRI is used in almost all fMRI and PET studies. Structural MRIs have a higher spatial resolution than functional images from MRI or PET. Thus, functional maps are overlaid onto the structural MRIs to identify the regional location of the functional activation. Whereas the coregistration is carried out on an individual basis, there is the need to average signals across subjects. Because each brain is different in absolute size and shape, the brain images have to be normalized before averaging. This is carried out by identifying landmarks on each brain, and then stretching and squeezing the images of different regions so the landmarks line up across brains. The human brain atlas that described the procedure most often used for normalization of brain structure is that of Talairach and Tournoux (1988). The simple basis of this method is to first align the brains in the same coordinate system. Once the coordinate system is established from these landmarks, the brain is divided into a proportional grid system by marking extreme points (e.g., the most anterior and posterior points). The compartments are then stretched and squeezed to scale the size to that of the brain in the atlas.

The brain in the Talairach and Tournoux (1988) atlas is, according to the authors, below average in size, which then requires considerable morphological squeezing for some subjects. To avoid this issue, some consortiums and institutes have used Talairach and Tournoux's grid strategy, but began to scale to a different "brain." For example, the Montreal Neurological Institutes created an average brain from 152 normal subjects, and the International

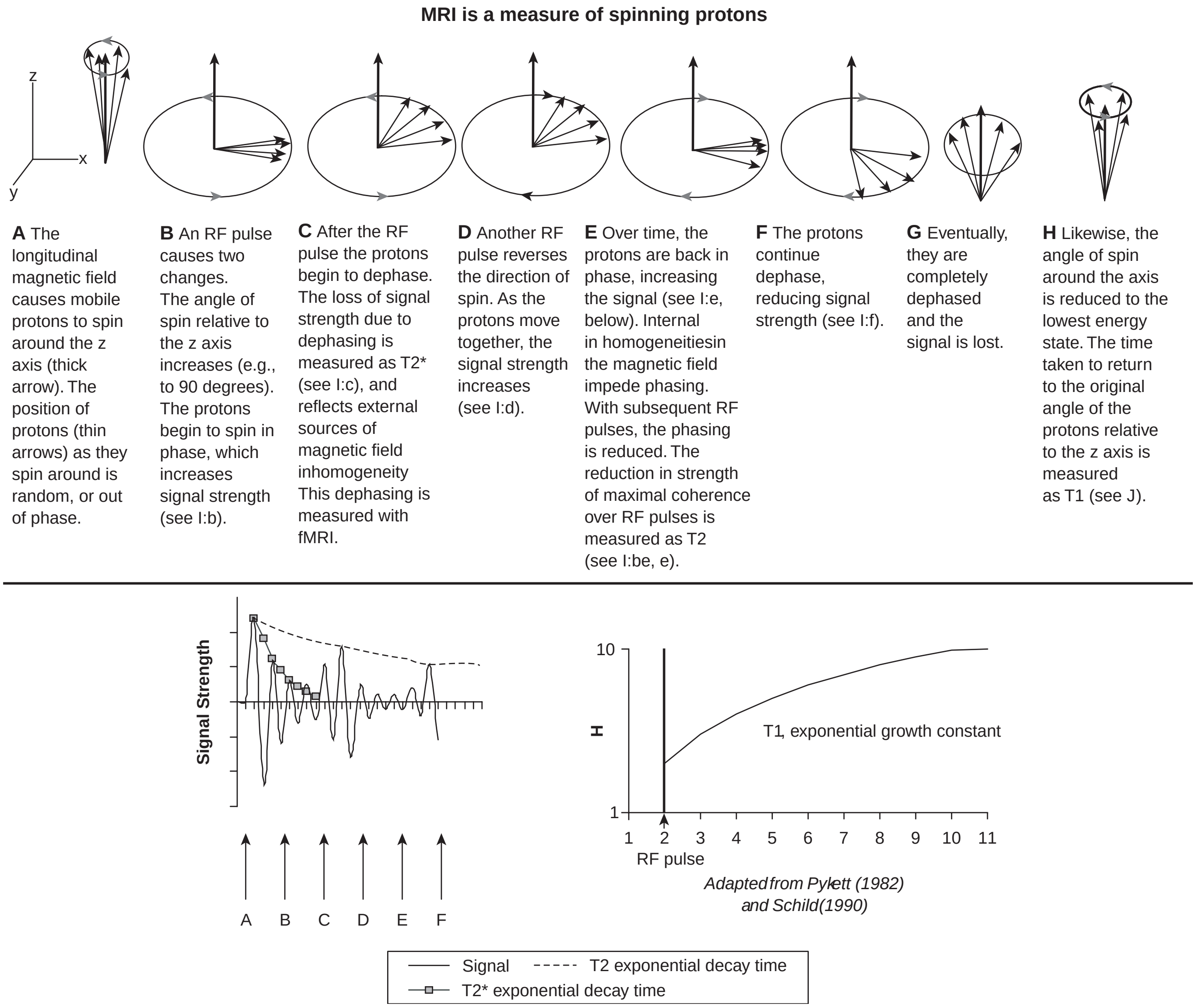

Figure 16.1 Structural MRI images and functional MRI images are based on measures of spinning protons. (A) A magnetic field is applied to create a longitudinal axis for protons (see bold arrow). Protons spin randomly around this axis. (B-F) depict the dephasing measured by the T_2* signal used for fMRI. The T_2 signal is used for structural MRI images and is a measure of the decrease in the peak strength after repeated rephrasing. (F-H) depict the reduction in the angle of the spin relative to the longitudinal axis. This T_1 signal can be used for structural MRI.

Consortium for Brain Mapping created an average brain from 452 normal subjects.

The idea of averaging a brain is somewhat suspect when we intuitively understand that there will be considerable variation in the sulcal and gyral boundaries across subjects, and that we are interested in localizing function to these structures. To compensate for this problem, brains can be "smoothed" before averaging. Thus, Talairach coordinates will not directly specify locations in sulci. Likewise, human brain areas are often referred to as "Brodmann's areas." But Brodmann defined his areas based on cytoarchitectonics, which do not necessarily correspond to sulcal and gyral boundaries (Schleicher, Amunts, Geyer, Morosan, & Zilles, 1999). With the highest resolution MRI methods, it may be possible to use layer size differences to delineate cortical areas (Walters et al., 2007); however, full cytoarchetectonic analysis requires analysis of layer size, cell size, and cell density. With these issues in mind, caution is required with regard to references to Brodmann's areas in human imaging studies.

FUNCTIONAL MRI

The development of more rapid imaging sequences for MRI was required before it was recognized that changes in blood flow and volume could be detected, and used to infer changes in neural activity. Because these more rapid imaging sequences provide a window into brain function, the general method is referred to as fMRI. There are several different MR pulse sequences that can be used to measure function. Blood oxygen level dependent (BOLD) imaging is by far the most commonly used, and will be discussed further. Briefly, the other methods, called perfusion-imaging techniques, are advantageous because they can quantify cerebral blood flow rather than reflect only relative changes in flow and volume. In these methods, an RF pulse is applied to one slice, and signal is measured from the image slice of interest above the slice receiving the pulse. This takes advantage of the fact that blood in arteries moves upward. The methods have the advantage of providing *quantitative* measures of cerebral blood flow (CBF) in absolute units, and a signal origin closer to the activated neurons than the BOLD signal. Because the image slice of interest is perfused with this endogenous tracer, these techniques are referred to as "perfusion imaging methods." They include the following variations: EPI STAR, FLAIR, FLOW, FAIR, and ASL.

The most commonly used functional imaging technique is BOLD. Iron in blood hemoglobin can be a source of magnetic susceptibility, and can cause a shortening of the T_2^* relaxation time (Tracey, 1998). Oxygenated hemoglobin (oxy-Hb) has a small magnetic susceptibility effect, but deoxygenated hemoglobin (deoxy-Hb) has a relatively large magnetic susceptibility effect. Thus, deoxy-Hb will cause faster dephasing and a shorter T_2^* relaxation time. Conversely, little deoxy-Hb (proportionally more oxy-Hb) will be related to slower dephasing and longer T_2^* relaxation time, or for our purposes, a greater signal intensity.

At any one time, at a baseline level of neural activity, the blood will have more oxy-Hb and less deoxy-Hb in arteries than in capillary beds, where oxygen is extracted, and in veins, which collect blood from capillaries. Having said that, the veins will have relatively shorter T_2^* relaxation times (i.e., faster dephasing) because of the magnetic susceptibility caused by the relatively high concentrations of deoxy-Hb (Tracey, 1998). But baseline neural activity and the differences in signals between arteries and veins are of little interest to psychologists.

Psychologists are interested in what happens when neural activity increases in relation to behavior or sensory input. As discussed before, increasing neural activity increases oxygen consumption, which, in turn, causes greater oxygen extraction. With greater extraction the proportion of paramagnetic deoxy-Hb with its high magnetic susceptibility is expected to increase in capillaries and veins. However, as noted earlier, that neural activity increases blood flow from arteries, and this would lead to an increase in availability of oxy-Hb. Taking no chances, this increase in blood flow *overcompensates* for the oxygen extraction, which in turn leads to the counterintuitive effect of the proportion of oxy-HbO_2 *increasing* in capillaries and veins when neural activity increases (see Figure 16.2). Conversely, the proportion of deoxy-Hb decreases, causing less magnetic susceptibility and *slower dephasing*. Altogether, this leads to a longer T_2^* relaxation time, or more simply a higher signal intensity. In summary, when neuronal activity increases in response to behavioral tasks, the relative decrease in deoxy-Hb has less influence on dephasing, and signal strength increases.

The vasculature can be compartmentalized into those vessels that carry blood into the brain (arteries), those that transport oxygen, glucose, and other chemicals into the brain tissue (capillaries), and those that passively carry blood out of tissue (veins). Arteries are capable of constricting or dilating, which influences the volume of blood in all three compartments. Capillaries are substantially smaller in diameter but higher in density than arteries, arterioles, venules, and veins. Deoxy-HB is expected to be present in capillaries and veins, rather than arteries (see Figure 16.2). At the magnetic field strengths, and with the pulse sequences typically employed for fMRI studies, the BOLD signal reflects primarily venous blood oxygenation, flow, and volume. Because veins are not immediately adjacent to the active neuronal population, the spatial resolution of the BOLD signal is limited by the distance between the site of neural activation and the veins carrying blood from that site. Thus, even when MRI methods have higher spatial resolution, that resolution is limited by the biological location of the signal being detected. The weight of signals from capillaries can be increased with high magnetic field strength and special pulse sequences (Goense & Logothetis, 2006). Nevertheless, these widely used methods have a spatial resolution that is sufficient to identify changes within regions and brain nuclei, and these are not trivial. For example, the visual representations in discrete areas of the retina can be mapped to specific locations within the primary visual cortex (DeYoe et al., 1996).

RELATION BETWEEN THE BOLD SIGNAL AND SYNAPTIC POTENTIALS

When interpreting the BOLD signal it is important to understand what aspect of neural activity is represented. Because the BOLD signal reflects changes in oxygenation, blood flow, and volume, it is expected, based on the previous discussion of physiology, to reflect synaptic signals. Researchers have confirmed this predication by the simultaneous collection of electrical recordings from synaptic field potentials, action potentials, and the BOLD signal in monkeys. Single recordings can measure multiple action potentials (multiple unit recordings) and synaptic potentials (local field potentials) simultaneously. Because action potentials are short in duration, they are signals with a high-frequency waveform. In contrast, synaptic potentials last longer and therefore produce lower-frequency waveforms. By filtering either high- or low-frequency signals each separate source can be compared to the BOLD signal. From simultaneous recordings, Logothetis and Wandell (2004) have shown that the synaptic potentials correlate

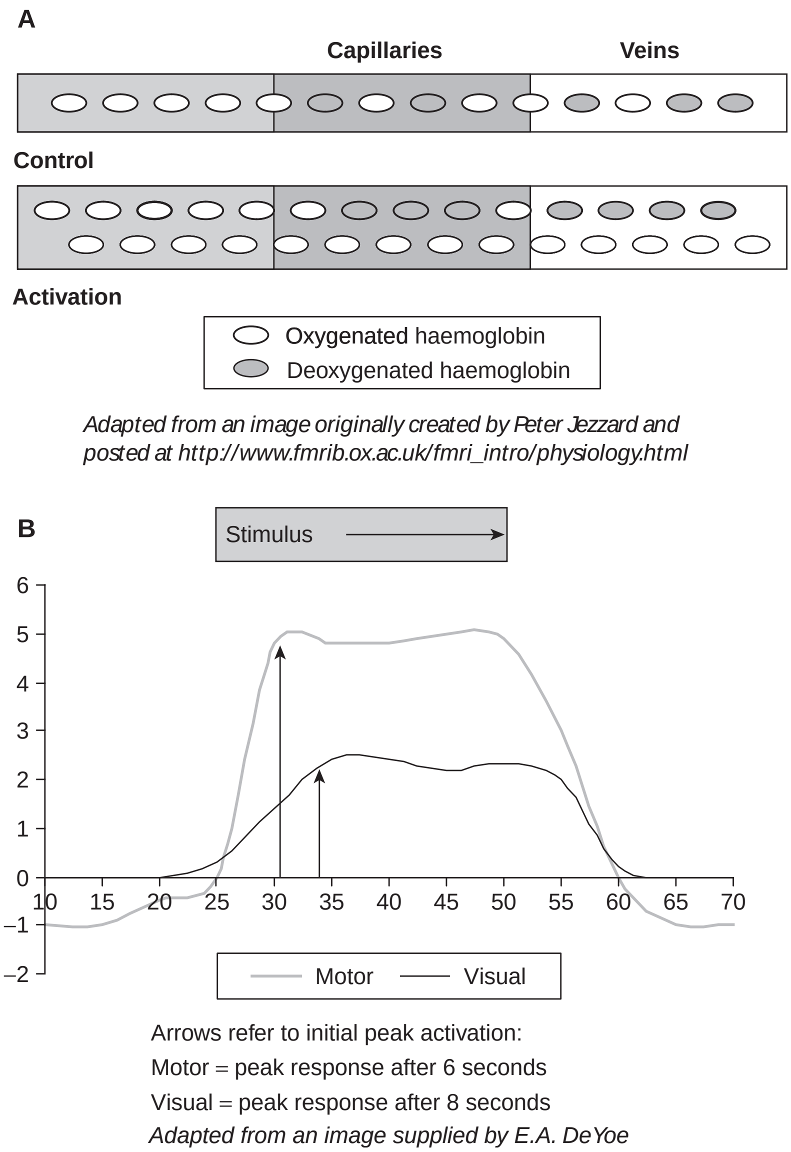

Figure 16.2 (A) FMRI reflects a blood oxygen level dependent (BOLD) signal called T_2*. The T_2* signal strength is influenced by inhomogeneities that cause dephasing of the spinning protons. Dephasing is caused in part by deoxygenated haemoglobin (Hb). During control conditions (upper schematic), oxygenated Hb is delivered by arteries to capillaries where oxygen is extracted. The deoxygenated Hb is then found in capillaries and veins where it causes dephasing. During neural activation (lower schematic), neural activity triggers dilation of arteries so that more oxygenated Hb is delivered to capillaries. Oxygen delivery exceeds the oxygen demands of increased neural activity producing a relatively greater ratio of oxygenated to deoxygenated Hb in the veins. The relatively lower levels of deoxygenated Hb reduces dephasing, leading to an increase in the T_2* signal strength. T_2* signal strength can be plotted over control periods and periods of neural activation (see B). The timing and strength of this BOLD signal can vary across brain regions. It is unclear whether the activation in the visual cortex (black) reaches an initial peak earlier than activation in the motor cortex (gray) because the visual stimulus is relatively more robust than the motor stimulus (flashing checkerboards versus touching the thumb to each finger successively) or because of differences in the hemodynamic responses across cortical regions.

better with the duration of the BOLD signal than with the duration of the action potentials. The action potential duration was brief and associated only with the onset of the stimulus, whether the duration was short or long, whereas the synaptic potentials were sustained and varied in duration, like both the stimulus and the BOLD signal.

The relation between the BOLD signal and synaptic signals is important to keep in mind as we compare studies using electrophysiological recordings and studies using fMRI. Invasive electrophysiological recordings often reflect single or multiple units, which reflect action potentials. From that traditional perspective of measuring neuronal activity, the BOLD signal may be a better reflection of action potentials from structures projecting into the structure imaged with fMRI. For example, the first CNS structure that receives visual information is the lateral geniculate nucleus of the thalamus (LGN), which in turn projects to the primary visual cortex. Action potentials in the LGN would produce synaptic potentials in the primary visual cortex that in turn drive the BOLD signal there. Thus, action potential recordings from the LGN rather than the primary visual cortex would be expected to correspond more closely to the BOLD signal. Unlike invasive recording methods, scalp electrode recordings reflect potentials from dendritic fields. Thus, EEG and fMRI data are compatible because both predominantly reflect postsynaptic potentials.

Originally, BOLD imaging was considered to detect only increases in signal, whereas studies using PET imaging reported decreases in glucose and oxygen uptake in regions. Eventually, negative signals were also observed with BOLD imaging. Simultaneous fMRI and electrical recordings suggest that decreases in the BOLD signal reflect a decrease in both local field potentials and multiple unit activity (Shmuel, Augath, Oeltermann, & Logothetis, 2006). However, the negative BOLD signal was 7 mm away from the stimulated positive BOLD signal, suggesting that the negative signal was not selectively coupled to the increase in neural activity.

TIMING OF THE BOLD SIGNAL

After increases in neural activity, the BOLD signal can take as long as 6 to 8 seconds to reach is maximal change (DeYoe, personal communication, see Figure 16.2). To take advantage of the maximal signal change, many researchers use a block design. In this design subjects perform a task repeatedly over a period of 20 to 30 seconds, and then have an equal length of time over which they perform a control task, or no task at all. Knowing the timing of signal change for brain areas, it is possible for the experimenter to design the expected signal change that should occur in regions active during task performance. This expected signal is then correlated with the observed brain signal measured with fMRI. The smallest unit from which a signal arises is called a “voxel,” which is the three-dimensional version of the more familiar word “pixel.” The expected signal can be correlated with the signal from every voxel measured, which can be a large number when the whole brain is imaged. With such a large number of voxels, and therefore statistical comparisons, substantial statistical correction for the number of comparisons must be carried out. Alternatively, some investigators apply a mask, which excludes voxels from regions that are not of interest or expected to be active during the task. This reduces the number of comparisons, or family wise error, so that lower correlation values, or t-scores, are necessary to reach significance.

The block design in fMRI was a natural transition from previous functional PET studies, which required long task performance intervals. The block design has obvious drawbacks, and did not allow for localization in tasks used for event-related potentials studies. Over time, the potential for event-related fMRI was recognized. It was already known that cerebral blood flow increased within two seconds of activation of sensory cortex (Dalkara, Irikura, Huang, Panahian, & Moskowitz, 1995; Woolsey et al., 1996), and the BOLD signal is capable of detecting this increase (Blamire et al., 1992). Eventually, detectable signal change was demonstrated after a stimulus lasting only 34 msec (Rosen, Buckner, & Dale, 1998). By averaging signals across trials, as carried out in event-related potential studies, the cancellation of baseline activity increased the detectability of the event-related BOLD signal change (Rosen et al., 1998).

Signals elicited by sensory stimulation or motor output have a higher signal-to-noise ratio than signals related to cognitive processing. However, event-related fMRI has more potential for the study of cognitive processing than sensory input or motor output. The possibility of using event-related fMRI for cognitive tasks was realized once it was demonstrated that a word generation task could evoke a fMRI signal change in both sensory and higher-order processing areas (Buckner et al., 1996). At high magnetic field strengths, averaging is no longer necessary; signal change related to a single trial can be detected.

DIFFUSION TENSOR IMAGING

In the human brain, white matter makes up nearly 50 percent of the volume in comparison to only 14 percent of volume in rodents (Goldberg & Ransom, 2003). It is now recognized that damage in white matter substantially contributes to dysfunction following stroke (Goldberg & Ransom, 2003). White matter in humans plays a key role in transmitting information across cortical areas as well as connecting cortical and subcortical areas. Accordingly, there is substantial interest in the use of MRI to study white matter and regional connectivity.

The direction orientation of axonal fiber tracts can be measured with diffusion tensor imaging. All MRI pulse sequences are sensitive to the spin of mobile protons, which are primarily protons in water. The movement of water in the brain is restricted by barriers such as membranes, organelles, proteins, cytoskeleton, and so on. Within cell bodies or dendritic fields, these barriers will have a relatively random orientation (perhaps with the exception of the cerebellar cortex). Within fiber tracts, however, there is a limited direction to the movement of water because axon diameters are relatively small and the parallel membranes form barriers to movement. Diffusion tensor imaging exploits this directional limitation for movement to identify the direction of axonal tracts. By applying gradient pulses in only one axis, the signal can detect diffusion relative to that axis. By applying diffusion gradients in at least six directions, it is possible to calculate the direction of diffusion for each voxel. (For an example of a DTI map, see http://www.dtiatlas.org/.) The reader should note that colors in DTI maps correspond to orientation of the fiber tracts.

Clinically, this method is very useful for detecting ischemia and disease-related changes in white matter (e.g., in MS or epilepsy). The method has also been applied to psychiatric populations. For example, using DTI, it was reported that schizophrenics had a shorter path length for the anterior limb of the internal capsule connecting to prefrontal cortex (Buchsbaum et al., 2006). This connectivity difference could help to explain the proposed disruption of the frontal-striatal-thalamic circuit. Although the most rigorous application of DTI involves efforts to map fiber tracts, which are particularly useful in combination with fMRI, it is important to remember that there is already a plethora of data on anatomical connectivity collected over the last century by neuroanatomists. The extent to which the relatively crude method of DTI adds to this knowledge remains to be seen. DTI is susceptible to a number of artifacts, leaving histological analysis the superior method (Miller, 2005). Histological analysis, however, is restricted to a single time point. DTI has the ability to measure path length *in vivo*, and therefore over time. Thus, anatomical changes in clinical populations could potentially be correlated with behavioral symptoms to aid in our further understanding disease processes and behavioral symptoms.

ARTIFACTS IN fMRI

It is beyond the scope of the present chapter to cover and explain all possible artifacts that can occur in fMRI studies. The user, properly trained, can prevent artifacts arising from improperly chosen fields of view, and can avoid regions of interest susceptible to chemical shifts that arise from differences in adjacent tissue. Likewise, artifacts arising from the movement of body fluids through the tissue can be compensated for faster image acquisition, and subjects can be trained not to move in the scanner. Other artifacts can be avoided only by frequent quality checks to detect circuit failure, failure in the shielding, and errors in the magnetic gradients used to localize the source of the signal (Hornak, 1996). Whereas service contracts provide repair for such problems, they are not yet ideal in practice because fMRI is not widely used for clinical purposes. Many psychologists have found that the success of their imaging research depends on the support of an MR physicist.

PET IMAGING

Before the advent of PET and fMRI, metabolic mapping studies were carried out by injecting glucose labeled with ^{14}C (2-deoxyglucose, 2-DG; Kadekaro et al., 1985). This form of glucose was not metabolized and therefore accumulated in cells. During task performance for 40 to 45 minutes, active neurons accumulated more 2-DG than inactive neurons. Localization of the radioactive label by film exposed to histological sections provides the ability to identify the active sites. As PET instrumentation was being developed, it became clear that imaging in humans was possible, but would require a safer tracer. The substitute was ^{18}F labeled glucose (2-fluoro-2-deoxy-D-glucose (^{18}F), FDG), developed at Brookhaven National Labs. Eventually, other safe tracers were developed. PET studies utilize radiolabeled FDG and radiolabeled oxygen, to provide a quantitative measure of cerebral metabolic rate of glucose (rCMRgl), regional cerebral metabolic rate of oxygen (rCMR02), regional cerebral blood flow (fCBF), and regional cerebral blood volume (rCBV; Roland, 1993). All of the tracer detection methods measure the isotope concentration in the brain in nCicc (nanoCuries per cc) as a function of time. These absolute measures are superior to the relative measures provided by BOLD fMRI. PET is not restricted to metabolic mapping. It is also possible to label transmitter analogues for the purpose of studying receptor and transporter density.

Despite its greater versatility, PET imaging is not as widely available as MRI. The requirement of injection or inhalation of radioactive tracers is a disadvantage, but the tracers have a relatively short half-life and are relatively safe. More restrictive is the need to synthesize tracers and use them quickly because of the short half-lives (2 to 120 minutes; Roland, 1993). Because of the short half-lives of the tracers, many facilities have their own cyclotron for local production of tracers. Some standard tracers can be shipped to a PET facility without a cyclotron.

PET is based on the emission of a positron from the tracer. The positron travels a short distance (0.2–7mm) and stops when it meets a negatively charged electron (Roland, 1993). The collision of the positron and electron causes them to "annihilate" and be converted to high-energy photons called *gamma rays*. The pair of gamma rays move in opposite directions, and hit sensors surrounding the head. The scanning device itself is like a scintillation counter divided into parts. When the gamma rays hit the scintillation material, they emit light, which is detected by a photomultiplier. By analyzing the coincident detection of gamma rays, it is possible to calculate the spatial location of the tracer within the brain, and therefore the tracer's location.

Tracer kinetic models describe the behavior of the tracer in different tissue compartments, such as the vasculature, extracellular space, cytosol, or subcellular space, or functional compartments such as precursor pools or metabolites.

LIMITS OF fMRI FOR THE STUDY OF BRAIN FUNCTION

Functional and structural imaging methods have opened windows into the localization of function in the human brain. The ability to correlate persistent hiccups to a small lesion in the medulla in a single patient with multiple sclerosis (Chang, W. H. Chen, Liu, Shih, & S. S. Chen, 1993) provides an example of the power of structural MRI for studying localization of function. To confirm brain activation during task performance through fMRI or PET provides the opportunity to search for convergence of data from studies of dysfunction and activation.

Whereas fMRI can localize active brain regions, it is limited in answering *how* information is processed in these areas. fMRI represents summed activity in dendritic fields. But the summed activity may not reflect more specifically how information is processed. For example, the contrast response function of the BOLD fMRI signal in the primary visual cortex approximates the perceptual contrast response function (unpublished data). But it seems unlikely that the elemental sensory features that contribute to the perceptual contrast response function would be recombined at this relatively early level visual processing. Data from electrophysiological recordings within the primary visual cortex suggest that within single cells the action potential frequency does not match the perceptual contrast response function. Instead, cells vary in their range of sensitivity. Some cells vary action potential frequency considerably over the limited range of 3 to 10 percent contrast, but not above or below. Others vary action potential frequency over the range of 10 to 30 percent contrast. Therefore, contrast in this early processing area is reflected by multiple channels (Albrecht & Hamilton, 1982). Albrecht and Hamilton proposed that the sum of activity over those multiple channels leads to the perceptual contrast response function, and the data from fMRI, which are essentially the summed input, support that hypothesis. The processing of contrast within single channels may be important for explaining why sensitivity to changes within a narrow range of contrast can be altered over time. Together this example reminds us that human imaging methods provide a summed response, and that there is a plethora of data collected in animals that can provide answers to the question of *how* information is processed. In summary, human functional imaging best answers questions about *where* active cells are located, but it is more limited in answering questions about *how* cells process information.

REFERENCES AND FURTHER READINGS

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology, 48*(1), 217–237.

Blamire, A. M., Ogawa, S., Ugurbil, K., Rothman, D., McCarthy, G., Ellermann, J. M., et al. (1992). Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. *Proceeding of the National Academy of Science, 89*(22), 11069–11073.

Borowsky, I. W., & Collins, R. C. (1989). Metabolic anatomy of brain: a comparison of regional capillary density, glucose metabolism, and enzyme activities. *Journal of Comparative Neurology, 288*(3), 401–413.

Buchsbaum, M. S., Schoenknecht, P., Torosjan, Y., Newmark, R., Chu, K. W., Mitelman, S., et al. (2006). Diffusion tensor imaging of frontal lobe white matter tracts in schizophrenia. *Annals of General Psychiatry, 5*, 19.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., et al. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceeding of the National Academy of Science, 93*(25), 14878–14883.

Chang, Y. Y., Chen, W. H., Liu, J. S., Shih, P. Y., & Chen, S. S. (1993). Intractable hiccup caused by medulla oblongata lesions. *Journal of the Formosan Medical Assocation, 92*(10), 926–928.

Clarke, D. D., & Sokoloff, L. (1999). Circulation and energy metabolism of the brain. In G. J. Siegel, B. W. Agranoff, R. W. Albers, S. K. Fisher, & M. D. Uhler (Eds.), *Basic Neurochemistry: Molecular, Cellular and Medical Aspects* (6th ed.). Philadelphia: Lippincott-Raven.

Dalkara, T., Irikura, K., Huang, Z., Panahian, N., & Moskowitz, M. A. (1995). Cerebrovascular responses under controlled and monitored physiological conditions in the anesthetized mouse. *Journal of Cerebral Blood Flow and Metabolism, 15*(4), 631–638.

Davatzikos, C., Shen, D., Gur, R. C., Wu, X., Liu, D., Fan, Y., et al. (2005). Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. *Archive of General Psychology, 62*(11), 1218–1227.

DeLisi, L. E., Sakuma, M., Maurizio, A. M., Relja, M., & Hoff, A. L. (2004). Cerebral ventricular change over the first 10 years after the onset of schizophrenia. *Psychiatry Research, 130*(1), 57–70.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceeding of the National Academy of Science, 93*(6), 2382–2386.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature, 427*(6972), 311–312.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Buchel, C., et al. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. *Journal of Neuroscience, 26*(23), 6314–6317.

Dunning, H. S., & Wolff, H. G. (1937). The relative vascularity of various parts of the central and peripheral nervous system of the cat and its relation to function. *Journal of Comparative Neurology, 67*, 433–450.

Erecinska, M., & Wilson, D. F. (1982). Regulation of cellular energy metabolism. *Journal of Membrane Biology, 70*(1), 1–14.

Gaser, C., & Schlaug, G. (2003). Gray matter differences between musicians and nonmusicians. *Annals of New York Academy of Science, 999*, 514–517.

Goense, J. B., & Logothetis, N. K. (2006). Laminar specificity in monkey V1 using high-resolution SE-fMRI. *Magnetic Resonance, 24*(4), 381–392.

Goldberg, M. P., & Ransom, B. R. (2003). New light on white matter. *Stroke, 34*(2), 330–332.

Hand, P. J., Greenberg, J. H., Sylvestro, A., Weller, L., & Reivich, M. (1979). A normal and developmentally local glucose utilization with natural stimulation of a single receptor organ. *Acta Neurologica Scandinavica Supplementum, 72, 60*, 46–47.

Hill, A. V., & Howarth, J. V. (1958). The initial heat production of stimulated nerve. *Proceedings of the Royal Society of London, Series B, Biological Sciences, 149*(935), 167–175.

Hornak, J. P. (1996, 2004). *The basics of MRI*. Retrieved December 15, 2006, from http://www.cis.rit.edu/htbooks/mri/index.html

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging*. Sunderland, MA: Sinauer Press.

Juliano, S. L., Hand, P. J., & Whitsel, B. L. (1981). Patterns of increased metabolic activity in somatosensory cortex of monkeys Macaca fascicularis, subjected to controlled cutaneous stimulation: a 2-deoxyglucose study. *Journal of Neurophysiology, 46*(6), 1260–1284.

Kadekaro, M., Crane, A. M., & Sokoloff, L. (1985). Differential effects of electrical stimulation of sciatic nerve on metabolic activity in spinal cord and dorsal root ganglion in the rat. *Proceeding of the National Academy of Science, 82*(17), 6010–6013.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (4th ed.). New York: McGraw-Hill.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology, 66*, 735–769.

Metea, M. R., & Newman, E. A. (2006). Glial cells dilate and constrict blood vessels: a mechanism of neurovascular coupling. *Journal of Neuroscience, 26*(11), 2862–2870.

Miller, N. R. (2005). Diffusion tensor imaging of the visual sensory pathway: are we there yet? *American Journal of Ophthalmology, 140*(5), 896–897.

Pitman, R. K., Gilbertson, M. W., Gurvits, T. V., May, F. S., Lasko, N. B., Metzger, L. J., et al. (2006). Clarifying the origin of biological abnormalities in PTSD through the study of identical twins discordant for combat exposure. *Annals of New York Academy of Science, 1071*, 242–254.

Pykett, I. L. (1982). NMR imaging in medicine. *Scientific American, 246*(5), 78–88.

Roland, P. E. (1993). *Brain activation*. New York: Wiley-Liss.

Rosen, B. R., Buckner, R. L., & Dale, A. M. (1998). Event-related functional MRI: past, present, and future. *Proceedings of the National Academy of Science, 95*(3), 773–780.

Roy, C. S., & Sherrington, C. S. (1890). On the regulation of the blood-supply of the brain. *Journal of Physiology, 11*(1-2), 85–158.

Schild, H. H. (1990). *MRI made easy…(well almost)*. Berlin: Schering.

Schleicher, A., Amunts, K., Geyer, S., Morosan, P., & Zilles, K. (1999). Observer-independent method for microstructural parcellation of cerebral cortex: A quantitative approach to cytoarchitectonics. *Neuroimage, 9*(1), 165–177.

Schultz, D. (1981). *A history of modern psychology* (3rd ed.). New York: Academic Press.

Shmuel, A., Augath, M., Oeltermann, A., & Logothetis, N. K. (2006). Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nature Neuroscience, 9*(4), 569–577.

Sokoloff, L., Reivich, M., Kennedy, C., DesRosiers, M. H., Patlak, C. S., Pettigrew, K. S., et al. (1977). [14C] deoxyglucose method for the measurement of local cerebral glucose utilization: Theory, procedure, and normal values in the conscious and anesthetized albino rat. *Journal of Neurochemistry, 28*, 897–916.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.

Tracey, I. (1998). *Brief introduction to FMRI—Physiology.* Retrieved from http://www.fmrib.ox.ac.uk/fmri_intro/physiology.html

Videbech, P., & Ravnkilde, B. (2004). Hippocampal volume and depression: a meta-analysis of MRI studies. *American Journal of Psychiatry, 161*(11), 1957–1966.

Vythilingam, M., Vermetten, E., Anderson, G. M., Luckenbaugh, D., Anderson, E. R., Snow, J., et al. (2004). Hippocampal volume, memory, and cortisol status in major depressive disorder: effects of treatment. *Biological Psychiatry, 56*(2), 101–112.

Walters, N. B., Eickhoff, S. B., Schleicher, A., Zilles, K., Amunts, K., Egan, G. F., et al. (2007). Observer-independent analysis of high-resolution MR images of the human cerebral cortex: In vivo delineation of cortical areas. *Human Brain Mapping, 28*(1), 1–8.

Wong-Riley, M. T. (1989). Cytochrome oxidase: an endogenous metabolic marker for neuronal activity. *Trends in Neuroscience, 12*(3), 94–101.

Woolsey, T. A., Rovainen, C. M., Cox, S. B., Henegar, M. H., Liang, G. E., Liu, D., et al. (1996). Neuronal units linked to microvascular modules in cerebral cortex: response elements for imaging the brain. *Cerebral Cortex, 6*(5), 647–660.

Zheng, D., LaMantia, A. S., & Purves, D. (1991). Specialized vascularization of the primate visual cortex. *Journal of Neuroscience, 11*(8), 2622–2629.

17

DRUGS AND BEHAVIOR

CATHY A. GROVER
Emporia State University

A drug is a chemical substance produced exogenously (outside of the body) that, when taken into the body, changes normal body functions. Psychologists are very interested in psychoactive drugs that change central nervous system (CNS; brain and spinal cord) activity, and thereby affect perception, thought, emotion, and behavior. Although people use many psychoactive drugs for acceptable medicinal reasons (see Chapter 18), this chapter focuses on those psychoactive drugs that people use primarily for recreational, nonmedicinal reasons (e.g., to feel good, be more alert, alter or avoid reality). An adult drinking alcohol to relax or smoking cigarettes to stop the jitters are examples of recreational use of licit (legal) drugs in this country, and smoking crack to feel euphoric or injecting heroin for the "rush" are examples of illicit (illegal) recreational drug use. Most of the information about how drugs make a person feel come from self-reports of licit and illicit users of drugs, whereas most of the data about how the body affects drugs and how drugs work in the brain comes from well-controlled experimental studies using nonhuman animals (see Chapter 15).

DRUG USE, MISUSE, ABUSE, AND ADDICTION

With pills to treat everything from the symptoms of the common cold to the positive symptoms of schizophrenia, drug use is prevalent in the United States, and pharmaceuticals are a multibillion-dollar industry. Nonetheless, society sends mixed messages about drug use, with commercials warning against the evils of illicit drug use and advertisements offering wonder treatments in a "purple pill." Drugs in and of themselves are not "evil," but every drug can be misused and abused. Misuse generally refers to the deviation from instructions on the label of over-the-counter drugs or the doctor's instructions for prescription drugs. For example, taking more or fewer pills per dose or day, not using the drug the full time course as prescribed (e.g., antibiotics), using the drug past the expiration date, using the drug with other drugs (e.g., alcohol with barbiturates), or sharing prescriptions with others without a doctor's permission are all forms of drug misuse. Although most of these acts may not seem very serious, they all can lead to very dangerous, even deadly, consequences (e.g., alcohol with other drugs, especially other CNS depressants). Drug abuse, which also can occur with licit and illicit drugs, refers here to use of a psychoactive substance to the extent that it produces some sort of physical, cognitive, behavioral, or social impairment. Keep in mind, however, that the public often thinks of drug abuse as specific to illicit drugs like methamphetamine, cocaine, heroin, and LSD (lysergic acid diethylamide), even though alcohol, for example, is a licit drug that people can abuse. What follows is an introduction to the use, misuse, and abuse of psychoactive drugs and their effects on behavior, beginning with how drugs enter the body and what happens to them once they do.

DRUG ADMINISTRATION, ABSORPTION, METABOLISM, AND EXCRETION

Pharmacokinetics is the study of how drugs are absorbed into the body, metabolized once in the body, and excreted from the body. The goal of drug absorption is for the drug to circulate in the blood, and more specifically for a psychoactive drug, the goal is for it to circulate in the brain. Administration for the purpose of absorption in blood and brain can take various forms depending on type of substance (lipid soluble vs. water soluble, gaseous vs. solid) and desired rate of absorption (rapid vs. slow, acute vs. continuous). Most humans administer drugs either orally (swallowed by mouth), sublingually (substance placed under the tongue), subcutaneously (injecting under the skin), intramuscularly (injecting into muscle tissue), intravenously (injecting directly into the bloodstream via a vein), transdermally (applied to outer layer of skin), intrarectally (using suppositories), intranasally (sniffed into the nostrils), or by inhalation (breathing gases and solids into the lungs). Intraperitoneal (into the peritoneal cavity), intraventricular (via a cannula into the ventricles of the brain), and intracranial (directly into a target area of the brain) injections are forms of administration used mostly in research with laboratory animals. Psychoactive drugs administered directly into the brain will have the most rapid effects because they will reach their CNS sites of action most quickly. Drugs administered through all other routes must be lipid-soluble in order to get through the formidable solid lipid barrier of the brain known as the blood brain barrier (BBB). Provided the psychoactive drugs administered directly into the bloodstream can pass the BBB, they will reach their CNS sites of action relatively quickly. Inhalation results in fast absorption into the bloodstream because gases and drugs in smoke (e.g., nicotine) are readily absorbed into the intricate network of capillaries that line the large surface area of the elaborately pocketed lungs. Although swallowing a pill is a simple, common method of drug administration, absorption is a tenuous process. Drugs taken orally must survive the harsh environment of the digestive system (e.g., stomach acids and digestive enzymes). Rates of absorption via other routes of administration are somewhere between those of inhalation and oral administration, depending somewhat on the availability of capillaries at the site of administration. Users of psychoactive drugs choose their favorite drug partially because of how quickly the drug exerts its psychoactive effects. For example, heroin is the preferred drug for some opiate addicts because it is more lipid soluble, is absorbed into the brain faster, and produces a faster, more intense “rush” than morphine does.

Before drugs can stop working in the body, they must be either broken down into other substances (metabolized) or removed from the body (excreted). Enzymes in the liver metabolize most of the psychoactive drugs described in this chapter into less lipid-soluble chemical products (metabolites). Some metabolites have their own effects on the body and brain. Several variables affect the rate of metabolism, including species, genetics, age, drug experience, and drug interactions. Regarding the latter, some drugs will inhibit or enhance the activity of enzymes responsible for metabolizing certain drugs—for example, SSRI-type antidepressants like fluoxetine inhibit some of the enzymes responsible for metabolizing codeine into the active analgesic morphine.

Subsequent to metabolism and recirculation in the blood, the kidneys excrete the more water-soluble metabolites from the body in urine, although there is excretion of small amounts of the drugs and their metabolites in exhaled breath, sweat, saliva, feces, and breast milk. Not surprisingly, urine drug tests are frequently used to determine the presence of metabolites some time after the metabolism of the original drug, rather than the presence of the original drug at time of administration.

HOW DRUGS WORK IN THE BRAIN

Pharmacodynamics is the study of how drugs work in the body. Psychoactive drugs work in the CNS. The brain (see Chapter 13) is made of supporting glial (fat) cells and excitable neurons. Neurons are responsible for the electrochemical transmission of information, enabling cells to communicate with one another. The structure and arrangement of neurons allows for the transmission, integration, storage, and interpretation of information received via sensory receptors, as well as the control of bodily organs and muscles. In other words, these specialized cells (neurons) are responsible for everything we think, feel, and do.

Psychoactive drugs work in the brain primarily by affecting neurotransmitter activity at the synapses (see Chapter 14). Neurons produce neurotransmitters that are stored in vesicles within the terminal buttons. When an action potential reaches the terminal buttons, vesicles release the neurotransmitter into the synaptic cleft. Neurotransmitter molecules then briefly bind to postsynaptic receptors causing ion channels to open, letting ions enter or exit, resulting in either excitatory or inhibitory postsynaptic potentials. Once released from the postsynaptic receptors, the neurotransmitter molecules undergo reuptake into the presynaptic terminal button or are destroyed by enzymes in the synaptic cleft. The effect of psychoactive drugs on synaptic activity can occur anywhere in the process of neurotransmitter production, storage, release, receptor binding, and reuptake or degradation.

More specifically, the administration of neurotransmitter precursors can increase the amount of neurotransmitter molecules available in the brain. For example, physicians prescribe L-DOPA, the precursor of dopamine, to patients with Parkinson’s disease in order to increase levels of dopamine in the brain. Other drugs can destroy or block the enzymes necessary for conversion of the precursors into neurotransmitters (e.g., p-chlorophenylalanine, PCPA, prevents the synthesis of serotonin). The vesicles that house neurotransmitters also can be the target site of drugs. Reserpine, occasionally used to treat high blood pressure, interferes with the transporter molecules that fill vesicles with neurotransmitters, thereby leaving the vesicles empty with no neurotransmitter available for release. A common target site of psychoactive drugs is the postsynaptic receptors where neurotransmitters bind. Drugs that bind to receptors and mimic the actions of a particular neurotransmitter are direct agonists (e.g., nicotine binds to nicotinic acetylcholinergic receptors). Drugs that bind to receptors without stimulating the receptor and prevent neurotransmitter molecules from occupying the receptor binding sites are direct antagonists (e.g., curare causes paralysis by binding to acetylcholinergic receptors). There are also drugs that work as agonists or antagonists by binding to sites other than where the neurotransmitter molecule binds (noncompetitive sites). Finally, drugs can affect what happens to neurotransmitters after they are released from their receptors by interfering with either the enzymes that break the neurotransmitters down (e.g., physostygmine deactivates the enzyme acetylcholinesterase, which breaks down acetylcholine) or reuptake mechanisms (e.g., cocaine deactivates the dopamine reuptake transporters).

Each type of neurotransmitter binds to a specific set of receptors that typically bear their own name (e.g., dopamine to dopaminergic receptors). The set of receptors belonging to a specific neurotransmitter can have quite different reactions to the neurotransmitter. Various drugs bind to receptor sites with different strengths, and may selectively bind to just one or a few receptor subtypes or nonselectively bind to multiple receptor subtypes. The selective serotonin reuptake inhibitors (SSRIs), used as antidepressants (see Chapters 18 and 89), bind specifically to receptor sites on the presynaptic serotonin reuptake pumps. Recent technological advances have allowed scientists to isolate the unique subtypes of receptor proteins, such that they can produce large quantities of each of the specific receptor proteins and then test the affinity of new drugs at each of the receptor subtypes. Drugs that bind to only a very specific receptor subtype or have greatly enhanced affinities for very specific subtypes of receptors will have fewer side effects as compared to drugs that bind less discriminately to an entire set of receptors. A good example is the drugs used to treat Parkinson's disease. Some of the older drugs (e.g., bromocriptine) that are structurally more similar to dopamine have more negative side effects than the newer developed drugs (e.g., ropinerole) that have more discriminate receptor affinity for D_3 than D_2 receptors.

Clearly, not all drugs are equal. Scientists who study drugs, pharmacologists, typically measure many participants' responses to doses so low that they cause no measurable effect to doses so high that they cease to cause any additional effect. They then plot the number (or percent) of participants who respond to the drug at each of the doses tested (dose response curve). The plot indicates the drug's *potency* (number of drug molecules required to elicit a given response), *efficacy* (the maximum effect of the drug, with additional amounts resulting in no increase in response), and *variability* (individual differences in responsiveness to the drug).

DRUG SAFETY AND TOXICITY

The federal Food and Drug Administration (FDA) has extremely specific guidelines in place for the testing of a new drug's effectiveness and safety (for more about the FDA's Office of Drug Safety go to http://www.fda.gov/cder/Offices/ODS/default.htm/). Safety refers to the drug's potential to produce everything from predictable, tolerable, unpleasant side effects to unpredictable, intolerable, severe toxicities. Unfortunately, all drugs have multiple effects, with some effects less desirable. Given that undesirable side effects are unavoidable, the goal is the most favorable combination of the most desired drug effects and the least unwanted side effects. The ED_{50} is the effective dose that produces the desired effect in 50 percent of the participants. The LD_{50} is the lethal dose that produces death in 50 percent of the subjects. Typically, the ED_{50} and LD_{50} are determined in several species and over many trials to reduce the risk of toxicity in humans. The greater the distance between the ED_{50} and $LD_{50,}$ the less the risk of drug-induced toxicity at beneficial dosages. The *margin of safety* is the ratio of LD_1 to ED_{99} (effective dose in 99 percent of the participants). Ratios of one or greater suggest greater safety. Be cautious, though—this margin of safety is under the best of controlled testing conditions, far from the circumstances under which many humans may take the drug (e.g., mixing drugs). One drug can alter the effects of another drug in many different ways. A second drug can have an additive effect, a synergistic effect (a greater effect than would be expected when just adding the two drugs), or an antagonistic effect (the second drug reduces or blocks the effect of the target drug). Even drugs not meant to have any effect (placebos) can influence a target drug's effects because of the user's expectations.

TOLERANCE, DEPENDENCE, AND WITHDRAWAL

Tolerance is the need to take increasing doses of a drug in order to achieve the same effects as previously achieved with lower doses. Likewise, when the same dose of drug has less and less of an effect with repeated administrations, tolerance has occurred. A good example is the tolerance that some people have for the caffeine in coffee. A novice coffee drinker may feel the stimulatory effects of coffee after a single cup of coffee containing about 100 mg of caffeine. After drinking coffee daily for a few weeks, it may take two or three cups of caffeinated coffee to feel that same excitation. There are several types of tolerance including metabolic tolerance, cellular tolerance, and behavioral tolerance. Metabolic tolerance occurs when, with repeated administrations of the drug, the body produces more and more metabolic enzymes, thereby speeding up the rate of metabolism of that drug. Thus, one must take more and more drug with each administration to maintain the same concentration of drug in the body as during previous episodes. Cellular tolerance is down regulation (reduction in numbers) of the receptors in the brain or reduced sensitivity of those receptors to the drug because of the continuous or repetitive presence of the drug. The result is the need for more drug in order to get the same level of effect in the brain. Behavioral tolerance involves learning (see Chapters 33 and 34). Behavioral tolerance can be observed in the presence of conditioned drug-taking cues and be absent in novel environments or situations. The drug serves as the unconditioned stimulus (US) and the drug effect as an unconditioned response (UR). Drug administering paraphernalia (e.g., white uniform, syringe and needle, bong and roach clips) and a specific location (e.g., doctor's office, nightclub, crack house) where the drug is administered can serve as conditioned stimuli (CSs) that, when paired with the drug (US), come to elicit conditioned responses (CRs) that are similar to the UR or opposite the UR (compensatory responses). For example, when Siegel (1975) gave rats morphine (US), they showed reduced sensitivity (UR analgesia) to heat applied to their paws, but with repetitive administrations of morphine in the presence of the same environmental cues (CS), the rats showed increased sensitivity (CR hyperalgesia) to those environmental cues.

A drug may develop different types of tolerance, and to all, some, or none of its effects. Some effects of a drug may even show acute tolerance, which occurs during a single administration of the drug. As a drug is absorbed into the blood, there is a gradual increase in the blood drug concentration, the ascending portion of the blood concentration curve. As long as drug administration has ceased, the blood concentration will eventually reach a peak level. When more metabolism than absorption is occurring, the concentration of drug in the blood begins to decline, depicted as the descending portion of the blood concentration. Acute tolerance is evident when, during a single administration, the measured effect is stronger on the ascending portion of the blood concentration curve than at the same concentration on the descending portion of the curve. Some effects of alcohol show acute tolerance in humans and rats. Finally, cross-tolerance is tolerance that occurs to one specific drug and subsequently tolerance occurs to the first administration of a different drug.

Drug dependence sometimes accompanies tolerance but does not require it. Dependence exists when a person must continue taking a drug in order to function normally and avoid the symptoms of withdrawal (physiological changes associated with the cessation of the drug). In other words, to determine dependence requires one to stop taking the drug. Physical dependence signifies that the body has adjusted physiologically to the repeated or continued presence of the drug. Removing the drug upsets the balance the body has established with the drug present, and results in often-unpleasant symptoms opposite those produced by the drug (e.g., heroin causes constipation whereas withdrawal from heroin causes diarrhea). Interestingly, long-term alcohol consumption can produce considerable tolerance without causing physical dependence; however, abstention from chronic alcoholism can cause life-threatening tremors, nausea, seizures, and delirium. Although the public tends to associate aversive withdrawal symptoms with drug "addiction" (e.g., chills, runny nose, fever, increased sensitivity to pain with opiate addiction), dependence and withdrawal are not unique to illicit drugs. Many legally prescribed and appropriately used therapeutic drugs result in dependence (e.g., SSRI-type antidepressants). It takes the body from several days to several weeks to readjust to the absence of a previously administered drug that has produced dependence. Thus, physiological dependence and withdrawal promote drug-taking behaviors, as people continue to take the drug, at least partially, to avoid the terrible effects of withdrawal. Several drugs that do not cause physiological dependence (e.g., cocaine, marijuana) do, however, produce psychological dependence. That is, an individual may be dependent on a drug for its pleasurable effects (i.e., positive reinforcement). Rats prefer to lever press for cocaine over food, even to the point of starvation.

SPECIFIC PSYCHOACTIVE DRUGS

Psychoactive agents are categorized a number of different ways. For example, drugs are categorized according to their chemical structure (e.g., amphetamines), whether they are legal or illegal (e.g., caffeine and nicotine vs. cocaine and amphetamines), how they affect the CNS (e.g., stimulants and depressants), and the type of behavioral, affective, and/or cognitive effects they produce (e.g., hallucinogens, analgesics). What follows is a description of a few of the most well-studied drugs with emphasis on the use of the drug; behavioral, cognitive, and mood-related effects of the drug; and the CNS mechanisms by which the drug produces its effects on behavior.

Stimulants

Stimulants produce behavioral excitation, increased motor activity, and increased alertness by enhancing excitation at neuronal synapses. The most commonly used stimulants include caffeine (a xanthine), nicotine, amphetamines, and cocaine. Many consider caffeine and nicotine to be "minor" stimulants and amphetamines and cocaine to be "major" stimulants. These stimulants have medicinal purposes, but most people are more familiar with their recreational uses. Stimulants vary greatly in the degree to which they affect behavior and in their potential for dependence and abuse.

Xanthines

Xanthines are a family of stimulants that includes caffeine, theobromine, and theophylline, the most widely used stimulants in the world. Caffeine is in many products (e.g., over-the-counter medications, baked goods, candy) but most commonly is associated with coffee and soft drinks. Tea contains caffeine, theophylline, and trace amounts of theobromine, and both caffeine and theobromine are in chocolate. Caffeine and theophylline are approximately equal with regard to stimulatory effects, but theobromine is only about one-tenth as strong as the other two.

How much caffeine is in coffee depends on the type of coffee bean (*coffea robusta* having twice the caffeine content of *coffee Arabica*) and how it is brewed (caffeine in a 5-ounce cup: instant about 60 mg, percolated about 85 mg, drip-brewed about 112 mg). Caffeine content in 12-ounce soft drinks ranges from about 38 mg (Diet Pepsi) to 54 mg (Mountain Dew), and as high as about 110 mg in special "energy" sodas (Red Bull). A 5-ounce cup of medium brewed black tea has about 60 mg of caffeine, and a strong brew of tea contains as much as 100 mg of caffeine. A 5-ounce cup of brewed tea contains a much smaller amount of theophylline (< 1 mg). A 1-ounce piece of milk chocolate contains 1 to 6 mg caffeine and about 40 mg of the 10 times less stimulating theobromine. There is 75 to 150 mg of xanthines in a cup of hot cocoa, and cocoa products contain enough caffeine and theobromine to affect behavior.

Orally consumed caffeine is absorbed in the stomach and mostly intestines, with peak blood levels occurring at 30 to 60 minutes. Caffeine easily crosses the blood brain and placenta barriers. Some foods, alcohol, smoking, hormones, age, and species affect the metabolism of caffeine. Xanthines are antagonists at adenosine A_1 and A_{2a} receptors, affecting the release of several neurotransmitters. When activated by adenosine, receptors located on presynaptic terminals inhibit spontaneous and stimulated neurotransmitter release. By blocking activation of adenosine receptors, xanthines lead to increased neurotransmitter release and increased excitation. At high concentrations, xanthines also block benzodiazepine receptors located on the GABA receptor complex, which may account for some of the increased anxiety after consumption of enormous amounts of coffee. Because outside of the CNS theophylline is particularly good at causing smooth muscles to relax, theophylline is useful therapeutically to dilate the bronchi of the lungs in the treatment of asthma.

Often people consume products containing moderate levels of caffeine because of their subjective experiences of increased alertness, improved attention, reduced fatigue, and more clear cognition. Experimental evidence suggests the most prominent effect of caffeine is enhancing performance of noncognitive tasks, like athletic and perceptual tasks, by reducing fatigue and boredom (see review by Weiss & Laties, 1962). Additionally, caffeine augments brainstem reflexes, enhances some visual processing, improves reaction time and self-reported alertness, reduces the detrimental effects of sleep deprivation on psychomotor performance, increases wakefulness, and produces insomnia.

Tolerance develops to some of the subjective effects of caffeine. Small and moderate, but not large, doses of caffeine appear to have reinforcing properties. Most people manage their caffeine intake, avoiding the anxiety, tremors, rapid breathing, and insomnia associated with high doses of caffeine. Within 12 to 24 hours of cessation, caffeine withdrawal often causes mild to severe headaches, drowsiness, muscle aches, and irritability, suggesting caffeine has some potential for producing limited physiological dependence. However, after reviewing caffeine studies, Nehlig (1999) concluded that caffeine does not affect the dopaminergic CNS centers for reward and motivation, as do cocaine and amphetamines.

Nicotine

Nicotine is one of the most-used psychoactive drugs in the world. The primary psychoactive active ingredient in tobacco is nicotine. According to the 2005 National Survey on Drug Use and Health (Department of Health and Human Services, 2006), about 71.5 million Americans (> 12 years old) used a tobacco product within the previous month. Only 30.6 percent of full-time college students ages 18 to 22 reported using in the previous month, as compared to 42.7 percent of same-aged part-time and noncollege students. Many of the toxic chemical compounds, other than nicotine, in tobacco products are the source of serious health problems (e.g., emphysema, chronic lung disease, cancer, cardiovascular disease) and death.

Nicotine is easily absorbed into the body. When inhaled, nicotine in cigarette smoke particles (tar) is quickly absorbed into the bloodstream via the capillaries lining the lungs. Smokers experience a sudden "rush" with that first cigarette of the day because the nicotine-saturated blood rapidly reaches the brain and crosses the BBB. Even though cigarettes contain about 0.5 to 2.0 mg of nicotine, smokers absorb only about 20 percent of that nicotine into blood. Smokers easily avoid nicotine toxicity by controlling the depth and rate of smoke inhalation. The liver metabolizes about 90 percent of the nicotine in the bloodstream before excretion. Urine tests measuring nicotine's major metabolite cotinine do not distinguish between tobacco use and environmental exposure.

Nicotine is an agonist at acetylcholinergic nicotinic receptors. Peripherally, nicotine's activation of receptors increases blood pressure, heart rate, and adrenal gland release of adrenaline. Nicotine activation of CNS nicotinic receptors located on presynaptic terminal buttons facilitates release of dopamine, acetylcholine, and glutamate throughout the brain. Physiological and psychological dependence of nicotine is due to nicotinic-induced release of dopamine from neurons projecting from the ventral tegmental area to forebrain regions (mesolimbic system) and prefrontal cortex (mesocortical system), brain areas responsible for reinforcement. Nicotine-induced release of acetylcholine is the likely cause of improved cognition and memory, as well as increased arousal. Increased glutamatergic activity due to nicotinic presynaptic facilitation contributes to enhanced memory of nicotine users.

Plenty of evidence exists regarding nicotine's facilitating effects on cognition and memory in humans and animals (for reviews see Levin, McClernon, & Rezvani, 2006; Levin & Simon, 1998). Individual differences in the cognitive effects of nicotine may be due to genetic variations in dopaminergic activity. Nicotine administered via a patch to adult carriers of the 957T allele (alters D_2 receptor binding in humans) impaired working verbal memory performance and reduced processing efficiency in brain regions important for phonological rehearsal (Jacobsen, Pugh, Mencl, & Gelernter, 2006). Additionally, nicotine stimulates activity in brain regions involved in attention, motivation, mood, motor activity, and arousal.

Tolerance appears to develop to the subjective mood effects of nicotine, but not to nicotine-induced changes in physiology or behavioral performance (for review see Perkins, 2002). However, most smokers do develop both physiological and psychological dependence on nicotine. Typically, withdrawal from cigarettes causes intense persistent cravings, irritability, apprehension, irritation, agitation, fidgeting, trouble concentrating, sleeplessness, and weight gain. Even people deprived of smoking just overnight report higher stress, irritability, and lower pleasure (e.g., Parrott & Garnham, 1998). Abstinence symptoms can last for several months, and many smokers find the cravings to be so intense that they relapse. It is common for smokers to quit smoking many times. Decreased activity in reward brain areas (e.g., dopaminergic mesolimbic system) that occurs during nicotine withdrawal may be responsible for the motivation of cravings, relapse, and continued smoking.

Amphetamines

In 1932 amphetamine, a synthetic drug similar in structure to ephedrine, was patented. That amphetamine is a potent dilator of nasal and bronchial passages easily administered as an inhalant made it a viable treatment for asthma in the early 1900s. During World War I and World War II, governments gave amphetamines to soldiers to prevent fatigue and improve mood. Subsequently, college students used amphetamines to stay awake studying for exams, and truck drivers for staying awake on cross-country hauls. It did not take long for word to spread that amphetamines (speed) caused euphoria, quickly making them an abused recreational drug. As Schedule II drugs, amphetamines have high potential for abuse and dependence, but also have accepted medicinal use with strict restrictions. Currently, treatments for narcolepsy and attention deficit hyperactivity disorder (ADHD) are accepted uses of amphetamines and amphetamine-like drugs (methylphenidate).

Amphetamines are a group of similarly structured synthetic chemicals that cause euphoria and behavioral stimulation. The *d* form of amphetamine is more potent than the 1 form. Administration is typically oral for current medicinal purposes, and inhalation or injection with a free-base form of methamphetamine (ice, crank) for a faster recreational "rush." Amphetamines easily cross the BBB and readily disperse throughout the brain. The liver metabolizes about 60 percent of methamphetamine, amphetamine being the major active metabolite, and then the kidneys excrete the metabolites and unchanged methamphetamine.

Amphetamines work both in the periphery and in the CNS. In the CNS, these drugs increase activity at synapses that release epinephrine, norepinephrine, and dopamine by either causing the neurotransmitters to leak out of their vesicles into the synaptic cleft and/or blocking reuptake into presynaptic terminal buttons. Recreational users typically prefer methamphetamine to other amphetamines because it has fewer unpleasant peripheral effects (e.g., increased heart rate, increased blood pressure, dry mouth, headaches) and stronger, longer-lasting CNS.

Amphetamines improve mood, decrease fatigue, increase vigilance, energize, impair ability to estimate time, and diminish the desire for food and drink. "Fen-Phen," a combination of fenfluramine and the amphetamine phentermine, was widely prescribed as an effective appetite suppressant in the 1990s, at least until it was removed from the market in late 1997 because of its association with heart valve problems and lung disease. Most of the performance-enhancing effects of amphetamines are limited to tasks that are routine, well-rehearsed, and well-practiced activities. Intravenously or intranasally administered high doses of amphetamine cause a "rush" of intense exhilaration and pleasure. The euphoria and strong reinforcing properties of amphetamines are due to increased dopamine activity in the mesolimbic system. Increased repetitive movements (stereotypy in laboratory rats) and behaviors (punding in humans) to the exclusion of eating, grooming, and sleeping are probably due to amphetamine stimulation in the nigrostriatal dopamine system. High acute doses and chronic use probably over stimulate the mesolimbic dopamine system, producing violently aggressive paranoia and amphetamine psychosis, delusions, hallucinations, and a split from reality. The sensation that insects are crawling under the skin (formication) may be the basis for the self-mutilation observed in laboratory animals. Long-term chronic use of methamphetamines is particular neurotoxic, leading to irreversible brain damage and psychosis.

Acute and chronic tolerance to amphetamines' desired effects of enhanced mood and euphoria occurs rapidly. The positively rewarding feelings associated with intravenously injected amphetamine, especially methamphetamine, leads to overwhelming psychological dependence. Physiological dependence on amphetamines is evident from the withdrawal symptoms of ravenous hunger, fatigue, lethargy, depression, and suicidal tendencies. Many of the characteristics of amphetamines are similar to those of cocaine.

Cocaine

For thousands of years, the natives of the South American Andes have increased endurance and stamina as they traveled the harsh mountain terrain by chewing the leaves of the coca plant. The plant became of interest to Europeans and Americans in the mid to late 1800s when entrepreneurs began adding the extract of the coca leaves to many products (e.g., wine, nerve tonics, home remedies, teas, and colas). In the 1860s Dr. W. S. Halstead discovered cocaine's local anesthetic properties. Because cocaine is readily absorbed in mucous membranes, it is still a local anesthetic of choice in some surgeries (e.g., nasal, esophageal). Currently, U.S. federal law categorizes cocaine as a Schedule II drug (high potential for abuse and dependence, but has currently accepted medicinal use with strict restrictions). In 2005 an estimated 2.4 million people were using cocaine, and about 2,400 persons per day used cocaine for the first time (Department of Health and Human Services, 2006).

Cocaine administration takes several forms, all with fairly quick but short-lived results (1 to 2 hours' duration). Users snort the powdered hydrochloride salt form of cocaine, and when they dissolve that in water, they can inject the drug. In the 1970s users developed a smokeable free-base form of cocaine by extracting the hydrochloride with the very volatile gas ether. The safer smokeable rock crystal crack cocaine forms when producers treat cocaine with baking soda and water. The crack user inhales the vapors as the rock heats and makes a crackling sound. When inhaled, cocaine is rapidly absorbed by capillaries in the lungs, whereas snorted cocaine hydrochloride is absorbed more slowly into mucous membranes. Cocaine readily crosses the BBB and quickly distributes throughout the brain, where it remains for as long as 8 hours. The major metabolite benzoylecgonine is inactive and, when excreted by the kidneys in urine, is detectable for 48 hours, even as long as 2 weeks in chronic cocaine users. Cometabolism of cocaine and alcohol produces the pharmacologically active, longer-lasting, and toxic metabolite cocaethylene.

Cocaine blocks presynaptic reuptake transporters for dopamine, epinephrine, norepinephrine, and serotonin. This blockade prolongs the presence of these neurotransmitters in the synapse, allowing the neurotransmitters to bind repetitively to postsynaptic receptors. Cocaine's enhancement of dopaminergic activity in the reward/reinforcement centers of the brain (e.g., the nucleus accumbens and other mesolimbic systems) is responsible for the highly addictive nature and powerful psychological dependence of cocaine. Serotonin receptors also play a role in the reinforcing effects of cocaine.

Cocaine is an extremely addictive psychostimulant that in low to moderate doses produces euphoria and increases alertness, mental acuity, self-consciousness, talkativeness, and motor behavior. Moderate to high doses cause more intense confusion, agitation, paranoia, restlessness, tremors, and seizures. Chronic use of cocaine produces impulsive and repetitive behavior. High-dose cocaine use can cause cocaine-induced psychosis characterized by extreme agitation and anxiety; exaggerated compulsive motor behaviors; delusions of paranoia and persecution; visual, auditory and tactile hallucinations; loss of touch with reality; and permanent brain damage. Medical risks associated with cocaine use include increased risk of cerebral ischemia, intracranial bleeding, heart attack and heart complications due to cocaine's vasoconstrictive properties, respiratory failure, strokes, seizures, and risks with snorting that include nasal lesions, perforations, bleeding, and infections.

Tolerance to cocaine's effects and physiological dependence to high doses of cocaine can occur. Regarding withdrawal syndrome, as the stimulatory CNS effects of cocaine subside, the user experiences depression, anxiety, lingering sleepiness, boredom, reduced motivation, and an intense craving for the drug. Much more powerful is the development of psychological dependence, because of cocaine's strong reinforcing properties, and therefore relapse.

Depressants

Depressants decrease CNS neuronal activity such that behavior is depressed, anxiety is lessened, and sedation and sleep are increased. This group of drugs includes barbiturates, benzodiazepines, some abused inhalants, and alcohol. Many of these drugs work at the GABA receptor complex, and all have potential for misuse, abuse, and dependence. The focus here is alcohol; see Chapters 18 amd 89 for discussion of the depressants used in pharmacotherapy.

Alcohol

Alcohol (ethanol) is a CNS depressant used throughout the world and history. In the United States, alcohol sales are an important part of the economy, with Americans spending over a hundred billion dollars annually on beer, wines, and distilled liquors. Based on alcohol sales in the Unites States, total ethanol consumption in 2004 was 377,002,000 gallons, including 4,368,000 gallons of ethanol and 97,065,000 gallons of beer (National Institute on Alcohol Abuse and Alcoholism, n.d.). Alcohol consumption in the United States costs in terms of increased risky behavior, injuries on the job, relational strain, and hospitalization. Chronic users develop vitamin deficiencies because alcohol is high in calories and not nutritious, and they are at risk for pancreatitis, chronic gastritis, gastric ulcers, stomach and intestinal cancers (alcohol is a gastric irritant), as well as death due to cirrhosis of the liver. Alcohol is involved in costly traffic-related injuries and fatalities. According to the National Highway Traffic Safety Administration, automobile crashes involving alcohol in 2000 cost the public almost $115 billion, with an estimated 513,000 people injured and 16,792 killed (Pacific Institute for Research and Evaluation, n.d.).

People typically absorb alcohol orally, and ethanol is easily absorbed via the gastrointestinal system. Generally, beers have 3.2 to 5 percent ethanol, wines 12 to 14 percent, and hard liquors (distilled spirits) 40 to 50 percent. An adult can metabolize the amount of ethanol contained in a 12-ounce, 3.2 percent beer, a 3 ½-ounce, 12 percent wine, or 1-ounce, 40 percent (80 proof) hard liquor in approximately one hour. The enzyme alcohol dehydrogenase metabolizes about 95 percent of the ethanol consumed into acetaldehyde at a constant rate of 0.25 ounces/hour, and that metabolizes to acetyl-coenzyme, which then converts to water and carbon dioxide. The other 5 percent is excreted unchanged mostly through breath (hence the use of Breathalyzers to estimate alcohol concentration). Women have less ethanol-metabolizing enzyme in their stomach wall and therefore absorb more ethanol than men do. Water and fat-soluble ethanol easily crosses the BBB and placental barriers. Ethanol diffuses rapidly throughout the brain, and ethanol concentrations in a fetal brain reach those of the alcohol-drinking mother.

Ethanol nonspecifically affects neuronal membranes and directly affects synaptic activity and ionic channels of several neurotransmitters. Ethanol dose dependently inhibits NMDA-type glutamate receptors (reduces postsynaptic excitation) and enhances inhibition produced by $GABA_A$ receptor-mediated influx of chloride ions (increases postsynaptic inhibition). Ethanol also induces synaptic release of opioids that trigger dopamine release in the brain reinforcement areas, explaining how the antagonist naltrexone reduces cravings for alcohol and relapse in alcohol-dependent persons attempting to abstain. Specific serotonin receptors ($5HT_2$, $5HT_3$) located in the nucleus accumbens may also be a site of ethanol action. Antagonists of those receptors reduce ethanol consumption in some persons with alcoholism. Additionally, ethanol leads to a decrease in the number of cannabinoid receptors (down-regulation) affecting the craving of alcohol.

Water excretion in urine increases (diuretic) as the blood alcohol concentration (BAC) rises, and water is retained (antidiuretic) as the BAC declines, causing swelling in the extremities. Although it makes the drinker feel warmer to the touch, ethanol actually causes hypothermia. Because ethanol causes blood vessels in skin to dilate, persons with white skin appear flushed. Behaviorally, ethanol has a biphasic effect, with low doses inhibiting inhibitions (disinhibition), and high doses depressing all behaviors. Alcohol reduces the latency to fall to sleep and inhibits REM sleep. Generally, low to moderate doses increase rate and tone of speech, impair visual perception, disrupt balance, worsen reaction time, exaggerate mood, reduce fear and apprehension (anxiolytic), and affect learning and memory in a state-dependent manner. However, there are huge individual differences in ethanol-induced effects because genetics, motivation, environment, experience with alcohol, and tolerance vary greatly. Chronic consumption of higher doses of alcohol can lead to memory storage problems, with heavy drinkers experiencing blackouts, periods during which they cannot remember events even though they were awake and active. Long-term heavy drinking can cause irreversible neuronal damage, producing dementia and severe cognitive deficits known as Korsakoff's Syndrome.

Each of the forms of tolerance can develop to some of the effects of ethanol (e.g., depression of REM) depending on pattern of drinking and amount consumed, with tolerance more apparent in regular and heavy drinkers. Physiological dependence is evident when withdrawal from ethanol results in agitation, confusion, tremors, cramps, sweating, nausea, and vomiting. Persons with severe alcoholism may experience delirium tremens (DTs) characterized by disorientation, hallucinations, and life-threatening seizures.

Opiates

The opiates, also known as narcotics, are a class of potent analgesics that have similar behavioral effects, including opium, opium extracts (e.g., morphine, codeine), several opiate derivatives (e.g., heroin), and several chemically unrelated synthetic opiates (e.g., methadone). Opium is harvested from the opium poppy, and has been used for centuries. In the 1800s, women and children alike ingested opium in everything from cure-alls to cough syrups. In the mid-1800s morphine as a medical analgesic increased with the invention of the hypodermic needle during the Civil War. When heroin was put on the market in the late 1890s it was considered a powerful and safe cough suppressant, and it was at least a decade before its full potential for abuse and dependence was realized. In the United States, opium, morphine, codeine, and methadone are all Schedule II drugs (high potential for abuse and dependence with acceptable medicinal uses with strict restrictions), whereas heroin is a Schedule I drug (high potential for abuse and dependence with no acceptable medicinal use). Although it is illegal in this country, it is widely used as a recreational drug. An estimated 108,000 persons age 12 and older used heroin for the first time in 2005 (Department of Health and Human Services, 2006).

Opiates are administered orally, as rectal suppositories, or as is most common with medicinal and recreational use, via injection. Morphine, because it is more water than fat soluble, crosses the BBB slowly, whereas more lipid soluble heroin crosses the BBB much more rapidly and produces a faster more intense "rush." Metabolism of morphine in the liver produces the metabolite morphine-6-glucuronide that is an even more potent analgesic. Heroin, which is three times more potent than morphine, metabolizes to monoacetylmorphine, which then converts to morphine. Urine tests detect the opiates as well as their metabolites, and therefore are not useful in determining the exact form of the drug used. Furthermore, poppy seeds and cough syrups contain ingredients that metabolize and test positive for opiate metabolites.

Exogenous opioids (produced outside of the body) bind to opioid receptors and mimic the actions of endogenous opioids (endorphins). Most of their analgesic effects are due to their presynaptic inhibition of pain-producing neurotransmitter release. There are Mu, Kappa, and Delta opiate receptors. Morphine is an agonist at Mu receptors located in several brain areas including the nucleus accumbens (addiction and abuse), thalamus, striatum, and brainstem (respiratory depression, nausea and vomiting), in the spinal cord (analgesia), and periphery. In the brain, Delta receptors in the nucleus accumbens and limbic system are involved in opioid-related emotional responses, and Kappa receptors may be Mu receptor–antagonists.

Opiates alter the perception of pain without affecting consciousness. They also produce a feeling of drowsiness, putting the person in a sort of mental fog that impairs cognitive processing. Opiates cause the user to feel carefree, content, and euphoric. Respiration becomes slow, shallow, and irregular, at therapeutic doses, and breathing may stop with high doses. Combined use of opiates and depressants can be particularly hazardous. Other physiological effects of opiate use include constricted pupils, histamine-induced red eyes and itching, lowered blood pressure, constipation, cough suppression, nausea and vomiting, and changes in the immune system.

Key features of frequent repetitive opiate use are tolerance to analgesia and euphoria, and cross-tolerance to other opiates. The severity of the withdrawal symptoms—anxiety, agitation, despair, irritability, physical pain, cramping, and diarrhea—depends on how long, how often, and how much of the opiate was used. Depression and intense cravings for the drug last for several months after the individual has stopped using heroin, for example.

Marijuana and Hashish

For centuries, people in many parts of the world have used some form of the *Cannabis sativa* hemp plant as a recreational drug. The plant grows as a weed in many parts of the United States, and is cultivated in other countries. Marijuana production consists of drying and shredding the leaves of the plant, whereas hashish is a dried form of the resin from the flowers of the plant. Marijuana and hashish are both Schedule I drugs in the United States. Each day in 2005, an estimated 6,000 persons, 59.1 percent under 18, used marijuana for the first time (Department of Health and Human Services, 2006). The main psychoactive substance in these products is delta-9-tetrahydocaanabinol (THC). Most marijuana has a THC content of about 4 to 8 percent.

Often THC is ingested via hand-rolled marijuana cigarettes called joints or reefers, but it is also consumed in cookies and brownies. Only about half of the THC in a marijuana cigarette is absorbed, but absorption via the capillaries in the lining of the lungs is very rapid. Peak blood levels of THC occur within 10 minutes of the beginning of smoking, and detectable blood levels continue for 12 hours after a single cigarette. Absorption and onset of effects after oral ingestion is slower, with peak THC blood levels not occurring for at least 2 hours. THC easily crosses the BBB and placenta barrier, and collects in fatty parts of the body. It slowly metabolizes into active 11-hydroxy-delta-9-THC, which then converts into inactive metabolites detectable in urine tests for as long as a month in heavy or chronic smokers.

The actions of THC in the brain were unknown until the 1990s identification of the weaker and shorter-lasting endogenous THC-like substance, anandamide, and the cannabinoid receptor. THC is an anandamide agonist that, when bound to cannabinoid receptors, inhibits the release of other neurotransmitters at those synapses, especially GABA. These g-protein-linked receptors are located mostly on presynaptic terminals and exist in large numbers throughout the brain, but not in the brainstem. That there are none of these receptors in the brainstem, where regulations of major life-support functions are controlled (e.g., heart rate, respiration), is probably why even high doses of marijuana are not likely lethal, although they can peripherally affect the cardiovascular and immune systems.

The psychoactive effects of THC are dependent upon type of administration, experience with the drug, environment, and expectations. Interestingly, the effects of marijuana on appetite and sexual behavior are culturally dependent, with Americans experiencing “the munchies” and Jamaicans decreased appetite, Americans enhanced sexual responsiveness and persons of India reduced sexual interest. Generally, people use marijuana because it has a mellowing, mildly euphoric effect. Other psychoactive effects include poorer attention, distorted perception of time and colors, altered auditory and gustatory perceptions, diminished anxiety, slowed reaction time, and impaired cognitive processing. Poor learning and memory are due to the numerous cannabinoid receptors in the hippocampus. Impaired balance and abnormal movements occur because of THC’s activation of receptors in the basal ganglia, and the cognitive effects are due to receptors in the cerebral cortex. Poor reaction time, decreased attention to peripheral visual stimuli, difficulty concentrating, and impairment of complex motor tasks under the influence of marijuana hinders driving ability. High doses of marijuana produce panic and anxiety, and extremely high doses produce additional disorientation, delusions, and hallucinations.

Tolerance, receptor-down regulation with repeated use, develops to THC. Withdrawal symptoms, which begin about 48 hours after the last marijuana use, include restlessness, irritability, despair, apprehension, difficulty sleeping, nausea, cramping, and poor appetite. Withdrawal symptoms are mild as compared to those that occur with most other drugs (e.g., alcohol, opiates) and occur for only about half of regular users. Thus, physiological dependence does occur for many, and most experience craving for marijuana when they stop using the drug.

The Hallucinogens LSD, MDMA, and PCP

Although the hallucinogens lysergic acid diethylamide (LSD), methylene-dioxy-methamphetamine (MDMA; Ecstasy), and phencyclidine (PCP; Angel Dust) affect very different neurotransmitter systems, they are grouped here because of their similar psychological effects at nontoxic doses. The commonality among hallucinogens (psychedelic drugs) is their ability to cause users to disconnect from reality and hallucinate. Although there are a large number of natural hallucinogenics, LSD, MDMA, and PCP are all synthetic drugs originally synthesized with hopes of medicinal value. Albert Hoffman accidentally experienced a "psychedelic trip" in 1943, and since then LSD has become one of the most widely known hallucinogens. PCP was used as an anesthetic prior to being taken off the market in 1965, when it made it to the streets as a recreational drug, and MDMA became a club drug in the 1960s. Estimates of first time users of hallucinogens for each year from 2000 to 2005 have been close to one million people age 12 and older per year, with about 615,000 first-time Ecstasy users in 2005 (Department of Health and Human Services, 2006).

Users ingest LSD orally, and the drug is absorbed in about 60 minutes with peak blood levels within 3 hours. LSD easily crosses both brain and placenta barriers. Metabolism takes place in the liver, where LSD is converted to 2-oxo-3phydroxy-LSD, and then excreted in urine. People use MDMA orally, snort it, smoke it, and inject it, with absorption being slowest with oral use. Most of MDMA metabolizes to 3,4-dihydroxymethamphetamine (DHMA), and urine tests detect metabolites for up to 5 days. People either take PCP orally (peak blood levels in 2 hours) or smoke the drug (peak blood levels in 15 minutes), and it is well absorbed into blood with both methods. Urine tests detect PCP for as long as a week after use.

LSD is a serotonergic receptor agonist, and most of the psychoactive effects of LSD are thought to be due to agonist actions at serotonin receptors in the pontine raphe (filters sensory stimuli). MDMA is structurally similar to but more potent than mescaline and metamphetamine, and stimulates release of both serotonin and dopamine in the CNS. Prolonged use of MDMA results in serotonergic neurotoxicity, and can lead to long-term verbal and visual memory loss. PCP blocks the ionic channels of glutamatergic-NMDA receptors, preventing calcium ions from entering into the dendrites of postsynaptic neurons when glutamate binds to these receptors, and blocking synaptic excitation.

Very small doses of LSD cause vivid visual hallucinations like colorful kaleidoscope lights and distorted images, lights to be heard and sounds to be seen, moods that oscillate from jubilation to dread, and frightening cognitions that surface. The LSD "trip" is often unpleasant, resulting in panic and confusion. A unique feature of LSD use is the infamous "flashback," a reoccurrence of the drug's effects without warning, even a long time after LSD use. At low doses MDMA has behavioral effects similar to those of methamphetamines, but at higher doses it has the psychoactive effects of LSD. Low doses of PCP produce agitation, dissociation from self and others, a "blank stare," and major alterations in mood and cognition. Higher doses of PCP can cause the user to have violent reactions to stimuli in his or her environment, along with analgesia and memory loss. Extremely high doses of PCP result in coma.

Tolerance and cross-tolerance develops to the psychological and physiological effects of most hallucinogens rather quickly, making it difficult to stay repetitively "high" on these drugs. Furthermore, there are few if any withdrawal symptoms and therefore little or no physiological dependence with LSD, MDMA, and PCP in most users.

SUMMARY

Psychoactive drugs change cognitions, emotions, and behavior. Drugs, per se, are neither good nor bad. People use psychoactive drugs for medicinal and recreational reasons. Regardless of the initial reason for using a drug, some people misuse and even abuse some psychoactive drugs because of the drugs' effects on mood, thought, and behavior. How people administer the drug (by ingestion, injection, inhalation, or absorption through the skin) affects how fast and intense the drug's effects are. The route of administration can affect how quickly the drug is absorbed into the bloodstream, how rapidly it is broken down, and how long it takes to be excreted from the body. The primary site of action for psychoactive drugs is the synapse of neurons in the brain. Often drugs work by either mimicking or blocking the actions of one or more neurotransmitters in the CNS. All drugs have multiple effects, and most are toxic at some dose. Dependence, withdrawal, and/or tolerance develop to some of the effects of most, but not all, psychoactive drugs.

Psychoactive drugs can be categorized many different ways. For example, by chemical structure, whether their use is legal or illegal, or the type of CNS or behavioral effects they produce. Stimulants, like cocaine and amphetamines, increase neuronal and behavioral activity. Depressants, like alcohol, reduce neuronal and behavioral activity. Opiates, some having legal uses (e.g., morphine) and others not (e.g., heroin), reduce pain and, in high enough doses, cause an addictive "rush." Low doses of marijuana have a mellowing, mildly euphoric effect, whereas very high doses can cause hallucinations. Drugs like LSD, MDMA, and PCP are all classified as hallucinogens because at even low doses they cause sensory and perceptual distortions. Although a great deal is known about how many psychoactive drugs act in the brain and affect behavior, researchers continue to identify the most effective pharmacological, cognitive, and behavioral treatments for persons who abuse these drugs.

REFERENCES AND FURTHER READINGS

Alford, C., Bhatti, J., Leigh, T., Jamieson, A., Hinmarch, I. (1996). Caffeine-induced sleep disruption: Effects on waking the following day and its reversal with an hypnotic. *Human Psychopharmacology: Clinical and Experimental, 11,* 185–198.

Baker, T. B., Japuntich, S. J., Hogle, J. M., McCarthy, D. E., & Curtin, J. J. (2006). Pharmacologic and behavioral withdrawal from addictive drugs. *Current Directions in Psychological Science, 15,* 232–236.

Department of Health and Human Services. (2006, September). Results from the 2005 National Survey on Drug Use and Health: National findings. Retrieved December 31, 2006, from http://www.oas.samhsa.gov/p0000016.htm#6tandard

Drummond, D. C. (2001). Theories of drug craving, ancient and modern. *Addiction, 96,* 33–46.

Griffiths, R. R., & Woodson, P. P. (1988). Caffeine dependence: A review of human and laboratory animal studies. *Psychopharmacology, 94,* 437–451.

Heishman, S. J., & Henningfield, J. E. (1992). Stimulus functions of caffeine in humans; Relation to dependence potential. *Neuroscience and Biobehavioral Reviews, 16,* 273–287.

Jacobsen, L. K., Pugh, K. R., Mencl, W. E., Gelernter, J. (2006). C957T polymorphism of dopamine D2 receptors gene modulates the effect of nicotine on working memory performance and cortical processing efficiency. *Psychopharmacology, 188,* 530–540.

Leri, F., Bruneau, J., & Stewart, J. (2003). Understanding polydrug use: Review of heroin and cocaine co-use. *Addiction, 98,* 7–22.

Levin, E. D., McClernon, F. J., & Rezvani, A. H. (2006). Nicotinic effects on cognitive function: Behavioral characterization, pharmacological specification, and anatomic localization. *Psychopharmacology, 184,* 523–539.

Levin, E. D., & Simon, B. B. (1998). Nicotinic acetylcoholine involvement in cognitive function in animals. *Psychopharmacology, 138,* 217–230.

Little, H. J., Stephens, D. N., Ripley, T. L., Borlikova, G., Duka, T., Schubert, M., et al. (2005). Alcohol withdrawal and conditioning. *Alcoholism: Clinical and Experimental Research, 29,* 453–464.

Malanga, C. J., & Kosofsky, B. E. (2003). Does drug abuse beget drug abuse? Behavioral analysis of addiction liability in animal models of prenatal drug exposure. *Developmental Brain Research, 147,* 47–57.

Martin, C. S., & Moss, H. B. (1993). Measurement of acute tolerance to alcohol in human subjects. *Alcoholism, Clinical and Experimental Research, 17,* 211–216.

Mumford, G. K., Evans, S. M., Kaminski, B. J., Preston, K., Sannerud, C. A., Siloverman, K., et al. (1994). Discriminative stimulus and subjective effects of theobromine and caffeine in humans. *Psychopharmacology, 115,* 1–8.

National Institute on Alcohol Abuse and Alcoholism. (n.d.). Volume beverage and ethanol consumption for States, census regions, and the United States, 1970–2004. Retrieved December 3, 2006, from http://www.niaaa.nih.gov/Resources/DatabaseResources/QuickFacts/AlcoholSales/consum02.htm

Nehlig, A. (1999). Are we dependent upon coffee and caffeine? A review on human and animal data. *Neuroscience and Biobehavioral Reviews, 23,* 563–576.

Pacific Institute for Research and Evaluation. (n.d.). Impaired driving in the United States. Retrieved December 3, 2006, from http://www.nhtsa.dot.gov/PEOPLE/injury/alcohol/impaired_driving_pg2/US.htm

Parrott, A. C., & Garnham, N. J. (1998). Comparative mood states and cognitive skills of cigarette smokers, deprived smokers, and nonsmokers. *Human Psychopharmacology: Clinical and Experimental, 13,* 367–376.

Patat, A., Rosenzweig, P., Enslen, M., Trocherie, S., Miget, N., Bozon, M. C., et al. (2000). Effects of new slow release formulation of caffeine on EEG, psychomotor and cognitive functions in sleep-deprived subjects. *Human Psychopharmacology: Clinical and Experimental, 15,* 153–170.

Perkins, K. A. (2002). Chronic tolerance to nicotine in humans and its relationship to tobacco dependence. *Nicotine and Tobacco Research, 4,* 405–422.

Pickens, R., & Thompson, T. (1968). Cocaine-reinforced behavior in rats: Effects of reinforcement magnitude and fixed-ratio size. *The Journal of Pharmacology and Experimental Therapeutics, 161,* 122–129.

Roth, M. E., Cosgove, K. P., & Carroll, M. E. (2004). Sex differences in the vulnerability to drug abuse: A review of preclinical studies. *Neuroscience and Biobehavioral Reviews, 28,* 533–546.

Siegel, S. (1975). Evidence from rats that morphine tolerance is a learned response. *Journal of Comparative and physiological psychology, 89,* 498–506.

Siegel, S. (2005). Drug tolerance, drug addiction, and drug anticipation. *Current Directions in Psychological Science, 14,* 296–300.

Siegel, S., & Ramos, B. M. C. (2002). Applying laboratory research: Drug anticipation and the treatment of drug addiction. *Experimental and Clinical Psychopharmacology, 10,* 162–183.

Walsh, J. K., Muehlbach, M. J., Humm, T. M., Dickins, Q. S., Sugerman, J. L., & Schweitzer, P. K. (1990). Effect of caffeine on physiological sleep tendency and ability to sustain wakefulness at night. *Psychopharmacology, 101,* 271–273.

Weiss, B., & Laties, V. G. (1962). Enhancement of human performance by caffeine and the amphetamines. *Pharmacologic Review, 14,* 1–36.

18

Behavioral Pharmacology

Wendy Donlin
University of North Carolina, Wilmington

Erin Rasmussen
Idaho State University

Behavioral pharmacology is an area of scientific study that bridges the fields of psychology and pharmacology, with subject matter covering any aspect of the interaction between drugs and behavior. Most often, behavioral pharmacologists study the influence of drugs on the entire behaving organism rather than focusing on the brain or periphery. The goal of behavioral pharmacology is to identify *behavioral mechanisms* involved in drug action (Branch, 2006).

Born out of learning theory, the field of behavioral pharmacology emerged from studies in two interesting, but quite diverse, research areas: (a) laboratory studies that showed that animals can seek out and administer certain drugs (called self-administration), and (b) the discovery and application of chlorpromazine as a psychoactive substance that effectively treats symptoms of schizophrenia. Scientists in behavioral pharmacology have since sought to combine the subject matter of pharmacology with the methodology and theory of operant behavior. From this strong scientific basis, behavior pharmacologists rigorously examine the functions and effects of drugs on behavior. Indeed, they discovered the most important tenets of behavioral pharmacology early within the development of the field: Drugs can function as reinforcers, as discriminative stimuli, and can change expressions of behavior.

SOME BACKGROUND ON OPERANT BEHAVIOR

To understand the principles of behavioral pharmacology, one must first understand a bit about operant behavior. The concept of operant behavior was developed in the early 1900s, but was refined by B. F. Skinner and his followers. Skinner found that most voluntary behavior has specific and predictable outcomes, or consequences, that affect its future occurrence. One type of behavior-consequence relationship, reinforcement, is the process in which behavior is strengthened, or made more likely to occur, by its consequences. For example, a common practice in operant psychology is to examine the relation between food, a common and important environmental stimulus, and behavior. In a laboratory, a slightly food-deprived rat is placed in a standard operant chamber and trained to press a lever that will produce a food pellet. If the frequency of the lever press increases with food pellet deliveries, then the food pellet serves as a reinforcer for lever pressing.

Conversely, punishment occurs when behavior produces an unpleasant stimulus, thereby reducing the likelihood that this behavior will occur again. For example, if a rat presses a lever for a food pellet, but the experimenter arranges the delivery of an additional environmental event, such as a mild shock, contingent on the lever press, lever pressing will likely decrease in frequency. The punisher (shock), then, overrides the reinforcing properties of food. In *Schedules of Reinforcement*, Ferster and Skinner (1957) described how arranging different types of reinforcing and punishing contingencies for behavior result in predictable changes in behavior, regardless of whether the specific behavior studied was a lever press, key peck, button press, or any other freely occurring behavior.

Ferster and Skinner also showed that organisms use environmental cues to predict the likely consequences of behavior. For example, in the presence of a particular stimulus, an animal may learn that certain behavior leads to a reinforcer. The relationship among the stimulus, behavior, and the result of that behavior is called the *three-term contingency*. It is represented by the following notation:

$$S^D: B \rightarrow S^{R+}$$

In this case, S^D refers to a discriminative stimulus, in whose presence a particular behavior, B, leads to positive reinforcement, S^{R+}. The discriminative stimulus can be any sort of environmental condition that an organism can perceive. In other words, the discriminative stimulus "occasions" responding and predicts that a specific behavior (B) will result in a specific consequence—in this case, the S^{R+}, or reinforcer. Consider a pigeon in an operant that has been taught to peck at a plastic disc in the presence of a green light. Pecking the disc when the light is illuminated green produces food reinforcement; pecking the disc under any other situation (e.g., a purple light) will not result in reinforcement. The pigeon, then, learns to peck the key only when the green light is illuminated. This process, called *discrimination*, is how most animals (including humans) learn to discern certain environmental situations from others.

The remainder of this chapter describes how drugs affect behavioral processes involved with the three-term contingency. More specifically, drugs may (a) affect behavior directly, (b) function as discriminative stimuli, and (c) act as reinforcers. Behavior pharmacologists have studied these behavioral processes rigorously in the laboratory, with the findings applied to real-world, drug-related human phenomena such as substance abuse. This chapter, then, begins by describing some principles of behavioral pharmacology discovered through basic research, and ends with a discussion of the application of these principles to pharmacotherapy and behavioral drug abuse treatment.

DIRECT BEHAVIORAL EFFECTS OF DRUGS

Administration of a psychoactive substance often directly affects behavior. For example, a drug may increase motor activity, inhibit coordination, or decrease the rate of behavior. Psychoactive drugs bind at receptor cites located throughout the central or peripheral nervous systems to exert pharmacological, physiological, and behavioral effects. As more of a drug is ingested, it exerts its influence on more receptor sites and thereby influences behavior.

In behavioral pharmacology, a change in behavior as a function of drug dose is represented by a dose-response curve. Figure 18.1 illustrates two hypothetical dose-response curves. Drug dose is represented along the x-axis as milligrams of the drug administered per kilograms of body weight. The y-axis shows response rate relative to baseline rates. Baseline rates of behavior (behavior without any drug present) are represented by the label "control," and saline-placebo condition is labeled "saline." Drug A shows no effect at 1 mg/kg, but it causes response rates to increase at 3 and 10 mg/kg and to decrease relative to baseline at the 30 mg/kg dose. The rise and fall of the curve represents a biphasic (two-phased) dose-response curve. Although this example is fictitious, this dose-response curve is representative of how stimulants such as amphetamine or cocaine affect behavior. Drug B shows a

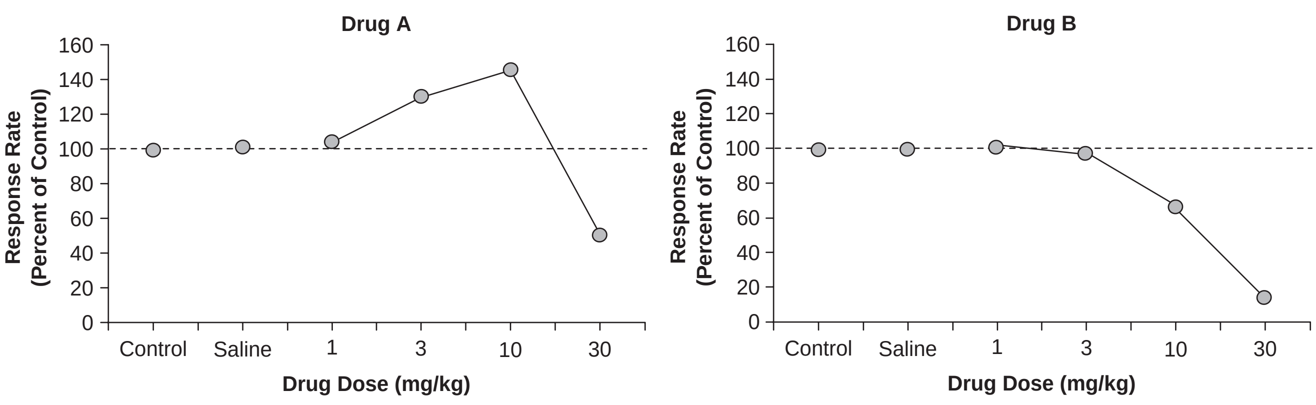

Figure 18.1 Response rate change relative to drug dose for Drug A (left panel) and Drug B (right panel). Response rate is shown as a percentage of baseline behavior for control, saline, and increasing drug doses. Drug A results in an increase in behavior, with a decrease in behavior only at the 30 mg/kg dose. Drug B has no effect on behavior at the two lowest doses, but decreases responding at the doses of 10 and 30 mg/kg.

different effect: Lower doses do not change behavior, but higher doses cause response rates to drop, relative to baseline. This curve, then, depicts a monophasic (one-phase) dose-response curve.

Rate Dependence

One of the most interesting discoveries in behavioral pharmacology is rate dependence: A drug's effect on behavior depends on the *baseline rate of behavior*. Dews (1955) first demonstrated this effect, and consequently launched the behavioral pharmacology movement. Pigeons pecked a disc under two different schedules of reinforcement: One schedule selected a high rate of disc pecking for food reinforcement; the other selected a low rate of disc pecking. Next, Dews administered pentobarbital, a depressant. The same dose of the drug had very different effects on behavior, depending on whether a high rate or a low rate of pecking occurred before drug administration. When the pigeon pecked at a low rate under baseline conditions, response rate decreased dramatically at one mg/kg dose. When the pigeon pecked at a high rate during baseline, this same dose increased response rate substantially. Although speculative, the phenomenon of rate dependence may have important implications for understanding why a stimulant such as Ritalin®, which is used to treat attention deficit disorder, may have calming effects for hyperactive individuals, but stimulating effects for typical individuals, who likely have lower baseline rates of behavior.

Behavioral History

As rate dependence exemplifies, it is often incorrect to make sweeping comments like "pentobarbital depresses behavior," "cocaine increases response rates," "morphine disrupts behavior," or "alcohol increases errors." Drug-behavior interactions are dynamic, and many variables affect how one influences the other. Behavioral history, or the types of experiences one has *before* taking a drug, plays a substantial part in the effects that a drug will have on behavior.

A classic study in behavioral pharmacology exemplifies that the manner in which discrimination is trained plays a role in the robustness of a drug's effect. Terrace (1966) trained one group of pigeons to discriminate a vertical line from a horizontal line by reinforcing pecking on the horizontal-line key while allowing these same pigeons to discover that pecking the vertical-line key would not produce reinforcement (called errors). With the other group of pigeons, he trained the discrimination differently—the introduction of the vertical line was so gradual that the pigeons responded very little, if at all, to it. (This procedure of fading in a second stimulus gradually is called *errorless learning*.) Next, Terrace introduced the drugs chlorpromazine and imiprimine (used to treat schizophrenia at the time) to both groups. Pigeons with a history of errorless training showed no signs of drug-induced disruption in their behavior: They maintained their discrimination of the vertical from the horizontal line. However, the pigeons with standard discrimination training (where errors were free to occur, and often did occur) showed disrupted behavior that was related to the dose of both drugs: the higher the dose of the drugs, the more disruption.

Terrace's (1966) study showed that drugs may affect the same behavior (discriminating line orientations) differently, depending on how the behavior is trained. Conversely, some drugs may affect *different* behaviors in a similar manner. For example, Urbain, Poling, Millam, and T. Thompson (1978) gave two groups of rats different reinforcement schedule histories to examine how behavioral history interacts with drug effects. They trained each group to press a lever at either high or low rates. After responding stabilized in each group, the experimenters placed both groups on a fixed-interval, 15-second (FI 15) schedule (the first response after 15 seconds elapses produces a reinforcer). The groups both maintained the differentiation of responding established by their previous training (either high or low rates), despite being under the same schedule. Then, the experimenters administered the drug d-amphetamine, a stimulant. Regardless of initial response rate, administering d-amphetamine produced responding at a constant, moderate level in both groups—it lowered the high response rates and increased the low response rates. This effect is termed "rate constancy."

As evidenced by the examples in this section, drugs can have both direct and indirect effects on behavior. Although psychoactive drugs primarily exert effects through interaction with central nervous system neurotransmitter systems, the context of drug administration is an important factor in influencing behavior.

DRUGS AS DISCRIMINATIVE STIMULI

When people or animals ingest a psychoactive drug, physiologically perceptible events may occur. For example, ingestion of over-the-counter cold medicine containing pseudoephedrine may occasion a person to report feeling "drowsy" or "nervous." Ingestion of a stimulant may occasion another person to report feeling "alert" or "hyper." Researchers can train animals and humans to discriminate whether they have been exposed to different types of drugs using these internal cues (called *interoreceptive stimuli*), much as they can be trained to discriminate among stimuli such as colors or line orientations in the external environment.

Researchers use signal detection procedures to test discriminability along sensory dimensions (e.g., a visual stimulus dimension, such as detail or acuity, or an auditory stimulus dimension, such as volume). The psychophysicist Gustav Fechner developed procedures to study the detection of "signals," or stimuli. He developed the concept of "absolute threshold," which refers to the lowest intensity at which a stimulus can be detected. He did so by repeatedly presenting stimuli and requiring subjects to indicate whether they detected that stimulus. Scientists can use these same procedures to test for the interoreceptive cues associated with ingestion of a drug, which is referred to as *drug discrimination*. For example, one may wish to determine the lowest dose at which a drug is detectable, whether an organism perceives two different drugs (e.g., morphine and heroin) as the same drug stimulus, and how organisms perceive combinations of different drugs (e.g., whether a small dose of ephedrine plus a small dose of caffeine feels like a large dose of amphetamine).

An example of a study in which researchers were interested in determining what drugs subjectively "feel" like morphine, Massey, McMillan, and Wessinger (1992) taught pigeons to discriminate a 5 mg/kg dose of morphine from a placebo (saline) by pecking one of two discs to produce food reinforcement. Pecking one disc (called the "saline disc") only resulted in access to grain with presession saline administration. Pecking the other disc (called the "morphine" disc) only resulted in access to grain with presession morphine administration. After the pigeons learned to discriminate the 5 mg/kg dose of morphine from saline, the experimenters administered other doses of morphine and doses of other drugs (d-amphetamine, pentobarbital, fentanyl, and MK-801) prior to experimental sessions. At very low doses of morphine, pigeons responded on the saline disc; they were unable to detect morphine administration, but could detect morphine at higher doses. Additionally, by administering drugs from different drug classes than morphine (an opiate), it is possible to interpret whether an animal can perceive differences among drugs. Fentanyl, also an opiate, occasioned responding on the morphine disc, but all of the other drugs resulted in responding on the saline disc. Thus, pigeons perceived fentanyl and morphine as similar but perceived that the other drugs are either different from morphine or not perceptible at all.

Humans can also discriminate opiates (Preston, Bigelow, Bickel, & Liebson, 1989). Former heroin addicts volunteered to discriminate three conditions: saline, hydromorphone, and pentazocine (both opioid agonists) from one another as "Drug A, B, or C," respectively, to earn money. After training, experimenters varied the doses of the substances in order to assess discriminability at different doses. Finally, Preston et al. administered drugs that partially mimic and partially block the effects of opioid agonists (butorphanol, nalbuphine, and buprenorphine) to test for generalization to hydromorphone and pentazocine. Not surprisingly, both hydromorphone and pentazocine resulted in greater discrimination with doses higher than the trained doses. The subjects also identified butorphanol as pentazocine at higher doses. Nalbuphine, however, was not identified as either opioid. Interestingly, the participants sometimes identified buprenorphine as pentazocine and sometimes identified it as hydromorphone, which suggests that buprenorphine has some properties similar to both drugs. Thus, drugs of similar neurotransmitter action can be discriminated from one another (e.g., hydromorphone vs. pentazocine), but some drugs are discriminated as the same drug (e.g., hydromorphone and butrophanol.)

Branch (2006) pointed out that discrimination procedures are useful at categorizing novel drugs as well as testing whether a novel drug has abuse liability. If a trained organism discriminates a novel drug similarly to known drugs of abuse, then further study on the abuse liability is warranted.

DRUGS AS REINFORCERS

A breakthrough for early behavioral pharmacology research involved the discovery that drugs could function as reinforcers. For example, in an early study, Spragg (1940) used food reinforcement to get a chimp to lie across a researcher's lap to receive a morphine injection. After regular morphine administrations, however, the chimp demonstrated that it "wanted" the injection by guiding the researcher to the location that injections took place, handing him the needle, and then bending over his knee. Later, Spragg gave the chimp two sticks that opened different boxes containing morphine or food. When deprived of either food or morphine, the chimp chose the corresponding box. Spragg's research produced the first evidence that a nonhuman animal would work to obtain a drug. Until this time, drug use/abuse was viewed as a uniquely human phenomenon, as well as a disease of personal character, and not a physiological or behavioral phenomenon.

Animal Research on Drug Self-Administration

Years after Spragg's findings, researchers began conducting laboratory-based studies to determine if drug self-administration could be studied with animals, and whether the findings could generalize to humans. Laboratory studies of animal self-administration are important because they allow researchers to manipulate variables and determine cause-and-effect relationships related to substance abuse that are not ethical to conduct with humans.

Headlee, Coppock, and Nichols (1955) devised one of the first controlled operant procedures for drug administration in animals. The experimenters pretreated rats with morphine, codeine, or saline for two weeks, and during this time also recorded the head movements of restrained rats. Then, they injected one of the drugs, but only if the rat held its head to a nonpreferred side longer than the preferred side during a 10-minute period. The rats shifted their head positioning to the nonpreferred side when receiving morphine or codeine, but not saline. These findings suggest that morphine and codeine were reinforcers for head movement.

Weeks (1962) simplified self-administration procedures by creating an automated apparatus and methodology to deliver a drug contingent upon lever pressing in an operant chamber. First, the researcher established physical dependence by repeatedly administering doses of morphine that elevated with each administration. At that point, experimenters connected a saddle to the rat, which held in place a cannula to inject a substance directly into the heart—a 10 mg/kg dose of morphine immediately contingent on programmed lever-press response requirements. Under this arrangement, the behavior of lever pressing increased, which indicated that morphine functioned as a reinforcer. When the experimenter reduced the dose of morphine to 3.2 mg/kg, response rates increased. Moreover, when they discontinued morphine injections, behavior eventually tapered to a cessation of responding after three hours, suggesting extinction of the drug reinforcer.

In a second experiment, Weeks (1962) increased response requirements for morphine, and in turn, the animals increased their responding to meet the response requirements for the drug reinforcer. When the experimenter used nalorphine, a morphine antagonist (which diminishes or blocks the effect of morphine) as a pretreatment, drug self-administration decreased dramatically. This study is important for several reasons: (a) it confirmed that morphine functioned as a reinforcer, (b) it showed predictable changes in behavior as a result of the type of reinforcement delivery; and finally (c) it showed that administering a morphine antagonist functioned as a condition to decrease the reinforcing efficacy of morphine, as evidenced by a decrease in morphine responding. In addition, the apparatus and procedures allowed easy self-administration in rats, which set the stage for studying drug self-administration today.

One drawback of Weeks' (1962) study as an animal model of drug use is that the experimenter initially created physical dependence to morphine by repeatedly administering morphine independent of behavior. Humans typically initiate drug use without physical dependence. The establishment of drug dependence as a necessary condition for drug dependence could have important implications for the generalizability of the data of this particular study to the problem of drug abuse. Recall that behavioral history is an important component that can change the effects of drugs on behavior. Studies since then, however, show that dependence is *not* a necessary or sufficient condition for drug self-administration.

For example, to determine whether dependence was necessary for self-administration, Pickens and T. Thompson (1968, as cited in Young & Herling, 1986) examined the reinforcing potential of cocaine without preexposure, while ruling out any extraneous possibilities for changes in behavior due to the physiological effects of cocaine such as motor function. Because cocaine acts as a stimulant, it could possibly produce increases in behavior that are related to increases in general motor activity, rather than motivation for cocaine reinforcement. An effective way to determine whether response increases are motivational in nature is to deliver the reinforcer without the lever-press contingency. If lever pressing continues when cocaine is "free," then it is assumed motivation for cocaine is not the reason for lever pressing. The experimenters initially delivered 0.5 ml of cocaine for lever pressing, and responding occurred at a steady rate. When the experimenters delivered cocaine freely (independently of lever pressing), lever pressing decreased dramatically. In another condition, responding ceased when the researchers administered saline instead of cocaine. This study convincingly linked drug self-administration in animals to its counterpart of human drug abuse. Animals initiated and maintained drug administrations, showing variability in these administrations with changing environmental conditions. Moreover, physical dependence was not a necessary condition for cocaine self-administration to occur.

For example, self-administration patterns in nonhumans often reflect self-administration patterns in humans. Deneau, Yanagita, and Seevers (1969) found that drug-naïve monkeys self-administered morphine, cocaine, d-amphetamine, pentobarbital, caffeine, and ethanol, all of which are drugs that humans self-administer. Interestingly, monkeys self-administered both morphine and cocaine continually until they caused their own deaths from overdose. They self-administered caffeine, but not consistently, and periods of abstinence occurred. Ethanol was not a reinforcer for all monkeys, and one monkey stopped administering it after a month of consistent administration. Most of the monkeys did not self-administer nalorphine, chlorpromazine, mescaline, or saline. The results from nonhuman animal research (e.g., see also Perry, Larson, German, Madden, & Carroll, 2005) are consistent with the general trends in human drug abuse, with humans abusing the same drug that animals self-administer.

The observation that an animal self-administers certain drugs does not always predict the magnitude of the abuse potential for the substance. Moreover, the rate of responding for a drug is not a reliable measure for abuse potential because each drug has different time courses of action, and pausing between self-administrations presumably varies due to action of the drug, its half-life, and drug dose available per administration. Thus, an additional method for quantifying the "addiction potential" of each substance is necessary.

Progressive-ratio schedules, in which the response ratio requirement for each successive reinforcer increases, allow for a measure of response strength not directly related to rate of responding. The response requirement for a progressive ratio schedule starts very small, with the first response often resulting in the reinforcer. The requirement for the next reinforcer increases in a systematic fashion, for example, from one lever press to five presses, and each subsequent requirement increases until the requirement is too high (in some cases, hundreds or thousands of responses for one reinforcer) to maintain responding (i.e., the animal no longer completes the response requirement for reinforcement). The highest completed ratio is designated as the "break point." Young and Herling (1986) reviewed studies of break points for drug reinforcers and concluded that cocaine maintains higher breakpoints than amphetamines, nicotine, methylphenidate, secobarbital, codeine, heroin, and pentazocine.

Drug Self-Administration as a Choice

Because there is strong evidence that drugs function as reinforcers, researchers have attempted to understand the choice to self-administer drugs over other nondrug-related behaviors, such as eating, or behaviors involved with “clean living.” The initial concept of choice behavior developed in the basic operant literature by way of the concurrent schedule of reinforcement. Ferster and Skinner (1957) described a concurrent schedule as two or more schedules of reinforcement that operate independently of each other at the same time. Each schedule may be associated with different reinforcers (e.g., food vs. drug), amounts of reinforcement (e.g., low dose vs. high dose of a drug), responses (e.g., lever pressing vs. wheel running), rates of reinforcement (e.g., high vs. low), and delays to the reinforcement (e.g., immediate vs. delayed). Consider the most commonly studied dependent variable—rate of reinforcement. Different schedules may yield different reinforcement rates under controlled conditions. For example, if responding at one lever produces five reinforcers per minute and responding on a second lever results in one reinforcer per minute, lever pressing under each schedule will result in “matching”: Behavioral allocation for the different schedules of reinforcement closely tracks the ratio of reinforcement available for the separate schedules of reinforcer delivery (e.g., Herrnstein, 1961). In this case, the organism will allocate five times more behavior to the “rich” lever compared to the “lean” lever. The study of concurrently available consequences for different behaviors closely approximates “real-world” behavior. Rarely, if ever, is someone faced with a circumstance in which only one behavior for one contingency can be emitted. The use of concurrent schedules is an ideal method for examining drug preferences.

Scientists have uncovered one drug-related principle with concurrent schedules: Higher doses of drugs generally are preferred over lower doses of drugs. Using concurrent schedules, Iglauer and Woods (1974) examined the effect of cocaine reinforcer magnitude on behavior allocation in rhesus monkeys. When the experimenters programmed large differences in cocaine doses, animals preferred the larger quantity of cocaine, regardless of how many deliveries occurred under each lever. However, when doses were equivalent, matching occurred based on the relative number of cocaine deliveries available on each lever.

Griffiths, Wurster, and Brady (1981) employed a different sort of task, a discrete-trial choice arrangement, to compare the choice between different magnitudes of food versus different magnitudes of heroin. Every three hours, baboons had an opportunity to respond for an injection of heroin or a food reinforcer. During baseline conditions, baboons chose heroin injections on 30 to 40 percent of the trials. The baboons often alternated food and heroin choices, rarely choosing heroin on two consecutive trials. Increasing the amount of heroin per dose, however, shifted preference away from the food and toward the drug.

One observation that arose from Griffiths et al.’s (1981) study and others like it is that self-administration of a drug depends not only on the magnitude of the drug reinforcer, but also on the value of other nondrug alternatives. Carroll (1985) provided evidence that indeed a nondrug alternative could alter drug self-administration. Saccharin and phencyclidine (PCP) were each available under a concurrent fixed-ratio schedule—conc FR 16 FR 16 (rats had to press one lever 16 times to produce access to reinforcers on that lever; the same reinforcement contingency also operated on the other lever). The experimenters varied three concentrations of saccharin and eight different magnitudes of PCP to test for preference. Increasing concentrations of saccharin moved preference toward the saccharin lever and away from the drug lever, but increasing doses of PCP resulted in more drug choices and fewer saccharin choices. Campbell, S. S. Thompson, and Carroll (1998) later showed that preference for PCP shifted to saccharin when the response requirement for saccharin was low and the response requirement for PCP was high. This research provided important evidence that the availability of alternative reinforcers and their respective magnitudes can alter the relative efficacy of a drug reinforcer.

The discovery that drugs act as reinforcers has important implications in understanding the conditions under which drugs are “abused.” Changing environmental conditions can lead to changes in drug self-administration and drug preference. Competing reinforcers can reduce drug use under certain arrangements, which is paramount in the treatment of drug abuse. Based on the basic behavioral pharmacology literature, researchers are developing behavioral treatments to control problem drug-related behaviors.

TREATMENT IMPLICATIONS OF BEHAVIORAL PHARMACOLOGY

It is useful to frame drug abuse as a problem with frequent drug self-administration. Based on studies throughout the history of behavioral pharmacology, drug self-administration can be reduced by lowering the rate of reinforcement (how frequently the animal receives a drug reinforcer), increasing the immediacy of drug reinforcer delivery, lowering the magnitude of the drug reinforcer, implementing a schedule of punishment involved with the drug reinforcer (also called *costs of self-administration*), increasing the response requirement of the drug reinforcer, increasing the availability of alternative nondrug reinforcers, and reinforcing behavior that is incompatible with drug self-administration (Bigelow & Silverman, 1999). Two types of treatment for drug abuse rely heavily on findings from behavioral pharmacology: pharmacotherapy and contingency management.

Pharmacotherapy

A drug that blocks the effects of a drug of abuse can often attenuate the reinforcing properties of the drug of abuse. For many drug classes, there are drugs that block a particular neurotransmitter system effect. For example, consider opiates. Naltrexone acts as an antagonist on the opiate system that can block the effects of opioid agonists, such as heroin or morphine. When administered together, agonists and antagonists work against each other, and the behavioral and physiological effects that result are a function of the potency and dose of each drug administered (Sullivan, Comer, & Nunes, 2006). Naltrexone, therefore, can be used as a treatment for opiate abuse because it blocks, among other things, the reinforcing properties of opiates (e.g., Preston et al., 1999). A drug similar to naltrexone, naloxone, also can be used to counteract opioid overdose immediately.

Disulfiram, or antabuse, blocks the degradation of alcohol by inhibiting the enzyme acetaldehyde dehydrogenase. This blockage causes a buildup of acetaldehyde in the blood, which results in unpleasant physical sensations similar to a "hangover" such as nausea and vomiting. Pretreatment with disulfiram is an effective treatment for alcohol abuse because ingesting alcohol after a dose will cause the individual to become ill—the aversive properties of drinking now override the reinforcing properties of the same behavior (e.g., Higgins, Budney, Bickel, Hughes, & Foerg, 1993). It should be noted, however, that because compliance in taking disulfiram is low, it is not used commonly today as a treatment for alcohol abuse.

Contingency Management

The field of applied behavior analysis, also based upon the tenets of operant psychology, developed in the late 1960s. The basic aspirations and goals of this field include applying the principles discovered through basic behavioral research to real-world human behavior. Researchers using this approach to behavioral change explored an array of subject matter, including the treatment of alcoholism.

One application of basic research with drug discrimination applies to humans. Lovibond and Caddy (1970) trained alcoholics to discriminate their blood alcohol concentration (BAC) from 0 to 0.08 percent by describing to the subjects what each blood level should feel like. When the subjects estimated their BACs, Breathalyzer results acted as feedback. (Breath alcohol concentration, which is measured by a Breathalyzer, closely approximates BAC.) Experimenters attached shock electrodes to the face and neck of 31 volunteers, allowed free access to alcohol, and informed the participants that they may drink with "impunity" as long as their BAC never reached 0.065 percent. If the participants' BACs were higher than this limit, they would receive a shock.

Experimenters varied the frequency of shock-punished trials, the duration of shocks, the intensity of shocks, and the delay of shocks through various phases of the experiment. Often, the experimenters encouraged family members to be present for conditioning procedures, and instructed them to support the alcoholic through social reinforcement for a low level of drinking. After several days of these conditioning procedures, the researchers encouraged participants to go about their everyday lives in a normal fashion. By drinking in a controlled fashion, 21 of the subjects had a positive outcome from the treatment, only rarely exceeding 0.07 percent BACs in normal settings. Three subjects improved considerably, only exceeding 0.07 percent BAC about once or twice a week. Four subjects only showed slight improvement, and the rest showed no change in drinking behavior. The authors reported no aversive reactions to alcohol, but indicated that subjects lost a desire to drink after a few glasses of alcohol.

In another study that took place in a specially equipped bar, 13 alcoholic volunteers learned to "drink using concurrent schedules" as a result of receiving finger shocks for exhibiting behavior typical of an alcoholic, such as gulping large amounts of alcohol or drinking straight liquor (Mills, Sobell, & Schaefer, 1971). The experimenters did not punish behavior associated with social drinking, such as drinking weaker drinks or taking small sips of alcohol. The type of drink ordered, the magnitude of alcohol consumed per drinking instance, and the progressive ordering of drinks determined the shock contingencies. After implementing shock, ordering and consuming straight alcoholic drinks decreased to virtually zero, the number of mixed drinks ordered remained fairly constant, and the total number of drinks consumed decreased. Sipping increased, and gulping decreased. The study, however, never tested for generality or maintenance of social drinking in other environments such as a real bar but did show that alcoholics could modify their drinking behavior without being abstinent.

Many researchers abandoned shock studies decades ago in favor of procedures that used positive reinforcement for moderate drinking or abstinence. Positive reinforcement techniques are as effective in modifying drinking behavior as the aversive conditioning techniques described. In one study (Bigelow, Cohen, Liebson, & Faillace, 1972), patients referred for treatment from an emergency room of a city hospital for alcoholism lived in a token economy-based research ward and participated in a variety of experiments examining their alcohol consumption. During control days, alcohol consumption was not restricted. On treatment days, subjects chose between a moderate level of alcohol consumption with access to an enriched environment (access to a television, pool table, interaction with other people, participation in group therapy, reading materials, allowance for visitors, and a regular diet) or a greater-than-5-ounce consumption of alcohol with access only to an impoverished environment (the participant was confined alone to a hospital bedroom with pureed food). On 76 percent of the treatment days in which the choice was in effect, the subjects chose to drink moderately and gain access to the enriched environment. The subjects did, however, consume the maximum amount of alcohol allowed under the enriched contingency.

Miller (1975) sought to reduce the amount of alcohol consumed by individuals in a population that exhibited high rates of public drunkenness offenses, unstable housing, and unstable employment. In the experimental group, Miller provided participants with goods and services such as housing, meals, employment, donations, and so on, until an instance of observed drunkenness or a Breathalyzer result revealed alcohol consumption. The control group did not have to refrain from drinking to have access to the same services. Miller reported that arrests decreased and employment or days worked increased. Blood alcohol levels were lower in the group with abstinence contingencies.

In another study, researchers used a behavioral contract to reduce drinking in one alcoholic, with the help of the alcoholic's spouse to enforce contingencies (Miller & Hersen, 1972). During baseline, the alcoholic consumed about seven to eight drinks a day. The husband and wife agreed to a contract in which the man could consume zero to three drinks a day without a consequence, but if drinking exceeded this level, he had to pay a $20 "fine" to his wife. The researchers also instructed the wife to refrain from negative verbal or nonverbal responses to her husband's drinking behavior, and she was to pay the same $20 fine to him for infringements on that agreement. After about three weeks, the husband's drinking consistently remained at an acceptable level, which maintained through a six-month follow-up. (Interestingly, the wife's nagging also decreased.)

Although the drinking behavior of alcoholics is modifiable using behavioral techniques, the general therapeutic community, most likely influenced by the early 20th century "temperance movement," regarded this proposition as absurd. Many people believe that alcoholics must completely abstain from drinking alcohol and that their drinking can never be stabilized at moderate levels. So deeply engrained is this idea of required abstinence that researchers have been slow to develop and popularize behavior modification techniques in training moderate drinking (Bigelow & Silverman, 1999; Marlatt & Witkiewitz, 2002). Fortunately, however, a community of therapists who treated addiction to substances other than alcohol embraced the same nonabstinence methods demonstrated to be useful in decreasing drinking.

One of the first descriptions of contingency management principles applied to controlling the behavior of drug abusers did not directly seek to reinforce drug abstinence; rather, it reinforced behaviors deemed important to therapeutic progress, such as getting out of bed before eight o'clock in the morning, attending meetings, performing work, and attending physical therapy (Melin & Gunnar, 1973). Drug abusers earned points for completing these activities, which moved subjects progressively from detoxification phases to rehabilitation phases, allowing desired privileges such as having a private bedroom, having visitors, and taking short excursions away from the ward. All behaviors that earned points increased in frequency, and the researchers concluded that access to privileges served as reinforcers for this population.

Contingency management is also successful in the treatment of heroin use with addicts receiving methadone maintenance treatment as outpatients. Hall, James, Burmaster, and Polk (1977) devised contracts with six different heroin addicts to change a specific problem behavior for each. Reinforcers ranged from take-home methadone (a less potent opiate used to treat heroin abuse), tickets to events, food, bus tokens, lunches, time off of probation, less frequent visits to probation officers, and access to special privileges. One subject increased his percentage of opiate-free urines (a marker of opiate abstinence) from 33 percent to 75 percent even up to six months after treatment. When rewarded for on-time arrival to work, a truck driver increased his on-time arrival from 66 percent to 100 percent. One client habitually tried to avoid treatment by arriving at the methadone clinic at strange hours to "swindle" her methadone out of a nurse not directly involved in her own treatment. When rewarded for showing up at the clinic during regular hours and receiving methadone from a specific individual, her frequency of arriving at regular hours increased from 13 percent to 100 percent compliance. Hall et al.'s (1977) study provided one of the first indications that treatment of outpatients with contingency management techniques for compliance of treatment goals for substance abuse was successful.

This early evidence of the utility of reinforcement procedures shaped a systematic development of modern-day contingency management. Different drugs of abuse have slightly different variations of contingencies employed while some aspects of this approach are uniform across treatments. Researchers seem universally to employ urinalysis as a measure of abstinence in abusers and deliver reinforcers for negative urine samples. The nature of the reinforcers, however, varies among different populations and individuals, the nature of the substance abuse, and the resources available as reinforcers.

The standard measurement tool for detecting drug abuse is urinalysis because the actual usage of drugs is often a covert behavior. Researchers cannot expect patients to provide reliable self-reports of drug use. Moreover, researchers test urine often enough to detect any instance of drug use. Testing need not occur each day for every drug, although it may be necessary for cocaine, codeine, or heroin because a small amount of these drugs used one day may be undetectable two days later. However, a researcher using too short of a testing interval may find more than one positive result from one instance of drug use; two to five days of abstinence may be necessary for a negative urine sample to be obtained in these conditions (Katz et al., 2002).

Researchers do not set drug abstinence as the only goal in the treatment of substance abuse. Contingency management is often supplemented with counseling, work therapies, self-help groups, relapse prevention strategies, and medical care. Compliance goals with these types of treatment programs often serve as behaviors to place under contingencies of reinforcement.

Experimenters use many different reinforcers for controlling behavior in these interventions—for example, vouchers that are exchangeable for goods and services. Researchers also use money, methadone, and services as common reinforcers. Money is an effective reinforcer that scientists have repeatedly shown competes favorably with a drug as a reinforcer (e.g., DeGrandpre, Bickel, Higgins, & Hughes, 1994). As the value of the monetary reward increases, drug preference weakens. However, with increasing delays to monetary reinforcement, preference for drug self-administration is more likely (e.g., Bickel & Madden, 1999). These discrete choice studies show larger amounts of money have greater reinforcing value than smaller amounts, similar to magnitude of reinforcer effects found in earlier animal studies.

Money seems to be the best reinforcer for drug abstinence because it does not have a limitation on magnitude. In addition, it has a low risk of satiation, and is effective for most people. A potential problem, however, is found in using money to reinforce drug abstinence—money can be used for future drug purchases (Silverman, Chutuape, Bigelow, & Stitzer, 1999). To avoid this potential problem, researchers use vouchers in its place.

Vouchers are exchangeable for many types of goods and services in most treatment programs. Often the value of a voucher starts quite low, for example, about $2. Researchers then increase the value of the voucher systematically with each successive negative urine sample or act of treatment compliance. This reinforcement arrangement is useful because it increases the period of abstinence with each reinforcer delivery. For example, the first day of reinforcement only requires a short abstinence period, but the next reinforcer requires the original time period plus the time to the next urine test. As shown in animal research, greater reinforcer magnitudes maintain responses that require more "work." Also embedded in this progressive reinforcer schedule is a response-cost contingency. If people choose to do drugs, they are not only losing the reinforcer for that particular day, they are losing the value of the difference between the original voucher value and the progressive value of the next potential voucher. Resetting the value of the reinforcer punishes drug-taking behavior. The progressive voucher system, then, is essentially a reinforcement of drug abstinence schedule with a response-cost punishment contingency for using drug that is simultaneously built into it.

Progressive reinforcement schedules are effective in maintaining drug abstinence in smokers. For example, Roll, Higgins, and Badger (1996) gave one group of smokers a $3 voucher for their first smoking-abstinent breath samples, with increments of 50 cents for each additional negative sample. They also gave a $10 bonus award for every third negative sample. Researchers gave a second group $9.80 for each negative sample, and gave a third group money that was equivalent to a paired individual in the progressive group (termed "yoking"), except these individuals did not need to be abstinent to receive the money. Both groups receiving contingent reinforcement exhibited more abstinence than the yoked control, but individuals in the progressive group exhibited the most abstinence of the three groups.

The data from contingency management studies nicely applies the fundamental principles of operant psychology to the area of substance abuse by assuming that drug taking is just one of many choice behaviors that have reinforcing value. Like many reinforcers, the reinforcing efficacy of drugs is not constant. Indeed, a variety of factors influence drug reinforcer efficacy, including the consequences for abstinence and the availability of alternatives.

SUMMARY

The field of behavioral pharmacology is more than 50 years old and has contributed immensely to our understanding of how drugs and behavior interact, especially in the area of substance abuse. Part of the reason for its success is the careful manner in which researchers characterize behavior-environment relations, which they study first in a laboratory setting and then apply to socially relevant problems. Researchers are better able to understand complex behavior related to drugs, such as drug use/abuse, when they consider it from the perspective of the three-term contingency. This conceptualization allows researchers to characterize drug effects in terms of their discriminative properties, reinforcing properties, and as modifiers of behavior. When scientists understand these behavioral mechanisms, the stage is set for effective treatments.

REFERENCES AND FURTHER READINGS

Bickel, W. K., & Madden, G. J. (1999). A comparison of measures of relative reinforcing efficacy and behavioral economics: Cigarettes and money in smokers. *Behavioural Pharmacology, 10*, 627–637.

Bickel, W., & Vuchinich, R. E. (2000). *Reframing health: Behavioral change with behavioral economics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bigelow, G., Cohen, M., Liebson, I., & Faillace, L. A. (1972). Abstinence or moderation? Choice by alcoholics. *Behaviour Research and Therapy, 10*, 209–214.

Bigelow, G., & Silverman, K. (1999). Theoretical and empirical foundations of contingency management treatments for drug abuse. In S. T. Higgins & K. Silverman (Eds.), *Motivating behavior change among illicit-drug abusers: Research on contingency management interventions* (pp. 15–31). Washington, DC: American Psychological Association.

Branch, M. N. (2006). How behavioral pharmacology informs behavioral science. *Journal of the Experimental Analysis of Behavior, 85*, 407–423.

Campbell, U. C., Thompson, S. S., & Carroll, M. E. (1998). Acquisition of oral phencyclidine (PCP) self-administration in rhesus monkeys: Effects of dose and an alternative non-drug reinforcer. *Psychopharmacology, 137*, 132–138.

Carroll, M. E. (1985). Concurrent phencyclidine and saccharin access: Presentation of an alternative reinforcer reduces

drug intake. *Journal of the Experimental Analysis of Behavior, 43*, 131–144.
DeGrandpre, R. J., Bickel, W. K., Higgins, S. T., & Hughes, J. R. (1994). A behavioral economic analysis of concurrently available money and cigarettes. *Journal of the Experimental Analysis of Behavior, 61*, 191–201.
Deneau, G., Yanagita, T., & Seevers, M. H. (1969). Self-administration of psychoactive drugs by the monkey: A measure of psychological dependence. *Psychopharmacologia, 16*, 30–48.
Dews, P. B. (1955). Studies on behavior I. Differential sensitivity to pentobarbital on pecking performance in pigeons depending on the schedule of reward. *Journal of Pharmacology and Experimental Therapeutics, 113*, 393–401.
Ferster, C., & Skinner, B. F. (1957). *Schedules of reinforcement*. Bloomington, IN: Prentice Hall.
Griffiths, R. R., Wurster, R. M., & Brady, J. V. (1981). Choice between food and heroin: Effects of morphine, naloxone, and secobarbital. *Journal of the Experimental Analysis of Behavior, 35*, 335–351.
Hall, S., James, L., Burmaster, S., & Polk, A. (1977). Contingency contracting as a therapeutic tool with methadone maintenance clients: Six single subject studies. *Behaviour Research and Therapy, 15*, 438–441.
Headlee, C. P. C., Coppock, H. W., & Nichols, J. R. (1955). Apparatus and technique involved in a laboratory method of detecting the addictiveness of drugs. *Journal of the American Pharmaceutical Association Scientific Edition, 44*, 229–231.
Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*, 267–272.
Higgins, S. T., Budney, A. J., Bickel, W. K., Hughes, J. R., & Foerg, F. (1993). Disulfiram therapy in patients abusing cocaine and alcohol. *American Journal of Psychiatry, 150*, 675–676.
Higgins, S. T., Heil, S. H., & Lussier, J. P. (2004). Clinical implications of reinforcement as a determinant of substance use disorders. *Annual Review of Psychology, 55*, 431–461.
Higgins, S. T., & Silverman, K. (1999). *Motivating behavior change among illicit-drug abusers: Research on contingency management interventions*. Washington, DC: American Psychological Association.
Iglauer, C., & Woods, J. H. (1974). Concurrent performances: Reinforcement by different doses of intravenous cocaine in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 22*, 179–196.
Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior, 77*, 129–146.
Katz, E. C., Robles-Sotelo, E., Correia, C. J., Silverman, K., Stitzer, M. L., & Bigelow, G. (2002). The brief abstinence test: Effects of continued incentive availability on cocaine abstinence. *Experimental and Clinical Psychopharmacology, 10*, 10–17.
Lovibond, S. H., & Caddy, G. (1970). Discriminated aversive control in the moderation of alcoholics' drinking behavior. *Behavior Therapy, 1*, 437–444.
Marlatt, G. A., & Witkiewitz, K. (2002). Harm reduction approaches to alcohol use: health promotion, prevention, and treatment. *Addictive Behaviors, 27*, 867–886.
Massey, B. W., McMillan, D. E., & Wessinger, W. D. (1992). Discriminative-stimulus control by morphine in the pigeon under a fixed-interval schedule of reinforcement. *Behavioural Pharmacology, 3*, 475–488.
Melin, G. L., & Gunnar, K. (1973). A contingency management program on a drug-free unit for intravenous amphetamine addicts. *Journal of Behavior Therapy & Experimental Psychiatry, 4*, 331–337.
Miller, P. M. (1975). A behavioral intervention program for chronic public drunkenness offenders. *Archives of General Psychiatry, 32*, 915–918.
Miller, P. M., & Hersen, M. (1972). Quantitative changes in alcohol consumption as a function of electrical aversive conditioning. *Journal of Clinical Psychology, 28*, 590–593.
Mills, K. C., Sobell, M. B., & Schaefer, H. H. (1971). Training social drinking as an alternative to abstinence for alcoholics. *Behavior Therapy, 2*, 18–27.
Peele, S. (1992). Alcoholism, politics, and bureaucracy: The consensus against controlled-drinking therapy in America. *Addictive Behaviors, 17*, 49–61.
Perry, J. L., Larson, E. B., German, J. P., Madden, G. J., & Carroll, M. E. (2005). Impulsivity (delay discounting) as a predictor of acquisition of IV cocaine self-administration in female rats. *Psychopharmacology, 178*, 193–201.
Preston, K. L., Bigelow, G. E., Bickel, W. K., & Liebson, I. A. (1989). Drug discrimination in human postaddicts: Agonist-antagonist opioids. *Journal of Pharmacology and Experimental Therapeutics, 250*, 184–196.
Preston, K. L., Silverman, K., Umbricht, A., DeJesus, A., Montoya, I. D., & Schuster, C. R. (1999). Improvement in naltrexone treatment compliance with contingency management. *Drug and Alcohol Dependence, 54*, 127–135.
Roll, J. M., Higgins, S. T., & Badger, G. J. (1996). An experimental comparison of three different schedules of reinforcement of drug abstinence using cigarette smoking as an exemplar. *Journal of Applied Behavior Analysis, 29*, 495–504.
Silverman, K., Chutuape, M. A., Bigelow, G. E., & Stitzer, M. L. (1999). Voucher-based reinforcement of cocaine abstinence in treatment-resistant methadone patients: effects of reinforcement magnitude. *Psychopharmacology, 146*, 128–138.
Spragg, S. D. S. (1940). Morphine addiction in chimpanzees. *Comparative Psychology Monographs, 115*, 1–132.
Sullivan, M. A., Comer, S. D., & Nunes, E. V. (2006). Pharmacology and clinical use of naltrexone. In E. Strain & M. Stitzer (Eds.), *The treatment of opioid dependence* (pp. 295–322). Baltimore: Johns Hopkins University Press.
Terrace, H. S. (1966). Errorless discrimination leaning in the pigeon: Effects of chlopromazine and imipramine. *Science, 140*, 318–319.
Urbain, C., Poling, J., Millam, J. D., & Thompson, T. (1978). d-amphetamine and fixed interval performance: Effects of operant history. *Journal of the Experimental Analysis of Behavior, 29*, 385–392.
Weeks, J. R. (1962). Experimental morphine addiction: Method for automatic intravenous injections in unrestrained rats. *Science, 138*, 143–144.
Young, A. M., & Herling, S. (1986). Drugs as reinforcers: Studies in laboratory animals. In S. R. Goldberg & I. P. Stolerman (Eds.), *Behavioral analysis of drug dependence* (pp. 9–67). Orlando, FL: Academic Press.

PART IV

Sensory Processes and Perception

19

Sensation

Christopher Koch
George Fox University

Psychology has roots in both philosophy and biology. This dual relation with philosophy and biology is probably most evident with sensation and perception. Philosophers have been interested in sensation and perception since our senses put us in contact with the external world; our perceptions are our representations of the external world. René Descartes believed that perception was among our innate abilities. John Locke and other empiricists, however, argued against innate abilities, instead suggesting that knowledge is based on experience obtained through the senses or by reflection on residual sensory stimulation. Locke separated our experiences into primary and secondary qualities. Primary qualities are inherent to an object (e.g., size) whereas secondary qualities are added to an object (e.g., color). George Berkeley disagreed with Locke's notion of primary and secondary qualities because he felt it created dualism between physical objects and mental perceptions. Berkeley believed only in perceptions. Because everything is perceived, the physical world is not necessary, therefore materialism is irrelevant. Berkeley, an ordained minister, further believed that the permanent existence of objects was due to God's perception of the world. An important issue associated with the empiricist view is whether we perceive the world veridically. If our knowledge is based on our perceptions of the world, do our perceptions accurately represent the world or are our perceptions subject to error? If our perceptions are subject to error, then our knowledge obtained from our senses is also subject to error (see Gendler & Hawthorne, 2006, for an overview of traditional and current issues in the philosophy of perception).

John Stuart Mills believed in the importance of experience but also added the idea of contiguity as a mechanism for acquiring knowledge and ideas through our senses. Contiguity refers to both space and time. Therefore, sensations that occur close together in space or time become associated with each other. Furthermore, associations can be made between present stimuli and images of stimuli hitherto removed. Alexander Bain added that associations between sensations result in neurological changes; however, the mental and biological operations take place in parallel.

These philosophical positions helped develop theories on how we learn, but they do not address the relation between how loud a sound is and how loud we perceive it to be, for example. Thus, psychophysics developed to determine the relation between a physical stimulus and the corresponding psychological experience (see Chapter 20 by Krantz in this volume). The three most influential people in the development of psychophysics were Ernst Weber, Gustav Fechner, and S. S. Stevens. Weber wrote *The Sense of Touch* in 1834 and proposed the concept of the just noticeable difference (JND). Fechner further contributed to psychophysics by expanding the work on JNDs and developed the method of limits, methods of adjustment, and method of constant stimuli as tools for assessing the relation between a physical stimulus and its mental counterpart. Although Fechner believed that there was a direct relation between a stimulus and the perception of it, he also believed that the physical and mental worlds operate in parallel with each other and do not interact (a position known as *psychophysical parallelism*). Stevens disagreed with the direct relation espoused by Fechner and instead argued that the physical and the psychological were related by an exponential function.

Wilhelm Wundt also played a significant role in establishing a scientific basis for examining perceptual processes. Although Wundt used systematic introspection, which was later called into question, he also used simple, discrimination, and choice reaction time tasks to examine our immediate conscious experience. Reaction times are perhaps the most widely used dependent variable in perception and cognitive research. Wundt used these methods to determine the elements of perception (elementism). Another interesting contribution of Wundt is that he believed that our conscious experience comprised not only sensations but also our affect. This belief highlights the fact that several processes are involved with influencing how we perceive the world and how we interact with it. However, Christian von Ehrenfels (1890), in a critique of Wundt, suggested that consciousness also contains Gestaltqualitaten (form qualities), aspects of experience that cannot be explained with simple associations between sensations. While at the University of Prague, von Ehrenfels influenced a young student, Max Wertheimer, who eventually developed Gestalt psychology.

Have you ever stopped at a traffic light, seen a car out of the corner of your eye, and then suddenly stepped harder on the brake pedal because you thought that you were moving? Something like this happened to Wertheimer on a train. The event caused him to realize that motion was perceived even though he was receiving sensory information to indicate that he was stationary. Thus, our perceptions are based on more than simple associations of sensory information. Instead, it is possible that a whole perception is actually different from the sensory parts it comprises. This realization led to the Gestalt psychology maxim—"The whole is greater than the sum of its parts"—and to a phenomenological approach for examining perceptual experience. As a result of this approach, Wertheimer proposed the Gestalt principles of perceptual organization (i.e., proximity, continuity, similarity, closure, simplicity, and figure-ground).

David Marr (1982) later proposed a computational theory of object perception. According to Marr, object recognition occurs through three stages of processing. First, a primal sketch or abstract representation is obtained from the retinal information, producing a blob-like image of contour locations and groupings. Additional processing of the primal sketch leads to a 2 ½-D sketch. This sketch yields an abstract description of the orientation and depth relations of the object surface to the viewer. Finally, a 3-D model is constructed that contains the perceptual object composed of a number of volumetric primitives (simple shapes like a cylinder) and the relationship between them. Thus, Marr believed that objects were perceived through a series of computational processes that render an image based on the relations between simple volumetric shapes. Similar to the use of volumetric shapes by Marr, Biederman (1987) proposed recognition by components (RBC) theory, in which simple geometric shapes, or geons (e.g., cylinder, cube, wedge), are combined to produce a mental representation of a physical object. The geons can vary across several dimensions to represent different parts of objects. For instance, if you place a medium-size snowball (sphere) on top of a large snowball, you have a snowman. If you place an object with a long, thin cylinder connected perpendicularly to a short, wide cylinder into the area of the snowman's mouth, you added a corncob pipe. A flattened cylinder (or disc) represents the button nose. Similarly, the two eyes made out of coal can be represented the same way.

Both Marr's computational theory and Biederman's RBC theory rely on contour information obtained through early stages of visual processing. One task that researchers have used extensively to examine the types of features available in early stages of visual processing is a search task. Search tasks require a participant to search for a target among a field of distractors. The number of distractors and features of the distractors are manipulated in the task. If you increase the number of distractors, you may increase the amount of time it takes to find the target because participants might engage in a serial search in which they process each item one at a time. However, if the number of distractors does not affect response time, it is likely that all of the items in the search array are being processed at the same time (parallel search). Serial searches take time and attentional resources. Parallel searches are fast and require no attention. Based on the results of search tasks, we know that features such as color, edges, shape, orientation, and motion are preattentitively extracted from a visual scene. According to feature integration theory, each of these features is placed into separate maps that preserve the spatial location of the features. Attention is then used to combine information from the different feature maps for a particular location. The combination of features leads to object perception. Feature integration theory highlights several important aspects of the perceptual process. First, perception takes place in stages. Second, our perceptions are influenced not only by the sensory information gathered from the environment but by other factors as well (e.g., attention and memory).

This overview of thought related to sensation and perception shows that philosophers and researchers attempt to address one key question: How do we derive an internal, mental representation of an external, physical world? Some approaches are phenomenologically based whereas others are computationally based. However, all theories of perception must be confined by the type of sensory information available for processing.

SENSATION AND PERCEPTION

The idea that we are information processors permeates many subfields within psychology (e.g., sensation and perception, cognition, and social psychology). An information-processing approach is fairly straightforward—there is input, processing, and an output or, put another way, we receive information, interpret it, and respond. Using the example of a computer, as I type, the computer scans for a key press. Any key press is very briefly stored in the keyboard buffer. However, when I type “a” an electric impulse is received and stored in the buffer as a binary code (not an “a”). Once the input is received, it is interpreted by the processor. The binary code of 01100001 then becomes an “a,” and I see “a” on the monitor in the document I am typing. Likewise, when light reaches the retina, the rods and cones undergo a bleaching process that produces a neural impulse that travels along the optic nerve, eventually reaching the lateral geniculate nucleus of the thalamus. The information received at the thalamus is not “what I see.” That information, therefore, is relayed to the visual cortex in the occipital lobe, where it undergoes further processing. The object is identified in the “what” pathway in the temporal lobe. A decision about what to do with the object is made in the prefrontal cortex and a response is generated in the motor cortex in the frontal lobe. Thus, within this information-processing account, sensation is the input, perception is the processing, and our response to the stimulus is our output. Sensation, therefore, is often described as the process by which stimulus energies from the environment are received and represented by our sense receptors. Perception, then, is the process by which the sensory information is interpreted into meaningful objects and events. Following this traditional separation of sensation and perception, if you are examining the sensitivity of the ear to particular frequencies of sound, you are studying sensation, and if you are examining the pitch of these sounds, you are studying perception.

It is also important to note that processing sensory information can occur in a bottom-up or top-down manner. Bottom-up processing proceeds from sensory information to higher-level processing (similar to the information-processing example just described). Top-down processing occurs when experience and expectations guide the interpretation of sensory information. Context effects are good illustrations of top-down processing. For instance, letters are easier to read when they make up a word than when they do not (Cattell, 1886). This phenomenon is referred to as the *word superiority effect*. Likewise, a degraded word can be recognized faster in the context of a sentence than when presented alone (*sentence superiority effect*; Perfetti, 1985). Similarly, portions of a line drawing can be recognized easier in context than out of context (Palmer, 1975). Facial features also appear to be recognized in the context of the whole face (*face superiority*). For example, the picture in the left frame of Figure 19.1 is easily recognizable. The picture in the middle, on the other hand, is unrecognizable even though all of the same features are present. Finally, in the right frame, the picture is recognizable again since the eyes and mouth are in the proper locations. However, turn the page upside down. The eyes and mouth are in the proper locations but not the proper orientations. In each of the preceding examples, anticipating the next word in a sentence or knowing the typical objects located at street corners or the features of a face influences how quickly the sensory information is processed.

As noted by Scharff in Chapter 27, Perception (this volume), the line between sensation and perception is sometimes fuzzy and not easy to delineate. In fact, sensation and perception are sometimes referred to as one continuous process in introductory psychology texts (e.g., Myers, 2007). For the purpose of this chapter, however, sensation includes the reception of energy from the environment (e.g., sound waves) and the process of changing that energy into a neural impulse (i.e., transduction) that ultimately travels to the appropriate area(s) in the brain. The other chapters in this section describe sensation and perception for our various senses.

Figure 19.1 This is an example of the face superiority effect and top-down processing. It is easy to recognize Britney Spears in the picture on the left. Even though all of the same features are present in the center picture, it is impossible to recognize the face. However, in the picture on the right, Britney Spears is recognizable because her eyes and mouth are in the correct locations even though they are inverted (turn the page upside down and look at this picture again).

MULTISENSORY PROCESSING

Although each sensory modality is typically researched in isolation of the other senses, it is easy to think of examples in which the senses interact. For example, when you have nasal congestion, you experience some impairment in your ability to smell things. You also cannot taste foods very well, since olfactory and gustatory senses are associated (e.g., Mozel, P. Smith, B. Smith, Sullivan, & Swender, 1969). Several types of sensory information interact with visual information. A common example is visual precedence. When you watch a movie, the action is occurring on the screen in front of you but the sound is coming from the speakers along the sidewalls. However, you attribute the source of the sound to the actors on the screen. The visual information takes precedence when attributing the source of the sound. Vision and touch interact as well. Graziano and Gross (1995) found bimodal neurons for touch and vision in the parietal lobe using monkeys. Specifically, they found that, with the eyes covered, particular neurons responded to the stroke of a cotton swab on the face. They also found that the same neurons responded when the eyes were uncovered and a cotton swab moved toward the face, indicating that the neurons not only responded to a tactile receptive field but to visual information associated with that receptive field as well. Another example deals with spatial processing. Although there is a "where" pathway for processing visual spatial information, spatial information is also processed in relation to our own position in space. Head-centric processing refers to processing objects in space in relation to our head. The semicircular canals function somewhat like a gyroscope, providing information about head orientation. The ocular muscles that move the eyeball receive excitatory and inhibitory information from the semicircular canals (Leigh & Zee, 1999). This input helps guide head-centric localization of objects.

Research in multisensory processing has been increasing in recent years and represents a growing area of research in sensory and perceptual processing. As this research continues, we will be able to gain a better understanding of how the senses interact with one another to produce the rich perceptual experiences we encounter from our environment.

SENSORY ADAPTATION

Our sensory systems are largely designed to detect changes in the environment. It is important to be able to detect changes because changes in the sensation (input) can signal a change in behavior (output). For instance, imagine you are driving down the street listening to music. Suddenly, you hear an ambulance siren. When you hear the siren you pull to the side of the road. The new sound of the siren signals a change in driving behavior. However, there are some stimuli that do not change. Usually, these constant sources of stimulation are not informative and our sensory systems adapt to them. Specifically, sensory adaptation (also neural adaptation) occurs when there is a change over time in the responsiveness of the sensory system to a constant stimulus. As a result, the firing rate of neurons decreases and will remain lower when the stimulus is presented again. An example of sensory adaptation is an afterimage. You have probably experienced a color afterimage demonstration. A common example is to look at a flag with alternating green and black stripes and a rectangular field of yellow in the upper left corner filled with black stars. When you look at this flag for approximately two minutes, your sensory system adapts to the green, black, and yellow colors. When you then look at a white background, you see alternating red and white stripes with a blue field in the upper left corner filled with white stars. This simple demonstration supports Hering's opponent-process theory in which color vision results from opposing responses to red-green and yellow-blue (also black-white). Adapting to green, for instance, reduces the neural response to green and increases the opposing response to red. Therefore, when you look at the white background, you see red where you had previously seen green. Sensory adaptation occurs in other areas of vision (e.g., motion aftereffect, spatial frequency) and in other senses. Researchers, therefore, can combine sensory adaptation with psychophysical methods to determine how the sensory system responds to physical characteristics of the environment (e.g., line orientation).

SENSORY MEMORY

Miller's (1956) well-known article "The Magical Number Seven, Plus or Minus Two" helped establish short-term memory as a fairly small memory store that held information for a limited amount of time (Brown, 1958: M. Peterson & L. R. Peterson, 1959). Sperling's (1960) use of the partial report method, however, provided evidence for a sensory memory store that could hold more information than short-term memory but for only a brief period of time. Atkinson and Shiffrin (1968), therefore, proposed a multistore model of memory that included sensory, short-term, and long-term stores. This model has been extremely influential in conceptualizing memory as an information-processing system and is still used today. Thus, the idea of sensory memory has been dealt with largely as an issue of memory and not necessarily sensation.

However, Phillips (1974) suggested that there are really two sensory-related stores. The first store lasts approximately 100 ms (for vision), has fairly high capacity, and is related to sensory processing. The second store holds information for up to several seconds, is limited in capacity, and is related to cognitive processing. In vision, this second type of store has become known as visual short-term memory (VSTM). It is hypothesized that sensory memory is needed to integrate sensory information obtained over time, albeit brief, as in a saccadic eye movement, or else sensory processing of the environment would start anew with each fixation rendering a new visual scene approximately every three to five seconds. Traditionally, there have been four characteristics of sensory memory that have distinguished it from other types of memory (Broadbent, 1958). First, the sensory memories are formed without attention. Second, sensory memory is modality-specific. Third, sensory memory is a high-resolution store, meaning that a greater level of detail is stored in sensory memory than is used in categorical memory, for example. Finally, information in sensory memory is lost within a brief period of time. Recent auditory research, however, has drawn into question these characteristics (Winkler & Cowan, 2005), leading some researchers to speculate about the degree to which memory stores can be separated (e.g., Nairne, 2002).

Although additional research is necessary to develop a more complete picture of sensory memory and how it interacts with and is different from other types of memory, sensory memory highlights an important point about sensory processing. As we interact with the environment, our senses provide us with information about what surrounds us. As we process that sensory information, we develop perceptions of the world. Although these perceptions develop quickly, they do take some time. Sensory memory, therefore, keeps information available for the additional processing necessary for us to understand that sensory information.

SUMMARY

Sensory processing provides a link between the external physical world and the internal neural world of the brain. Sense receptors convert energy from the physical world into neural impulses that are processed in various cortical areas as basic components of a visual object, a sound, and so on. Sensory information is held briefly in sensory memory for additional processing, leading to our perceptions of the world. Sense receptors also reduce the amount of information we process since they are only receptive to a particular range of physical stimuli. Additionally, we can adapt to sensory information that is not informative. Finally, although sensory systems can be examined independently, our sensory systems are integrated to give us complex perceptual experiences about the world we interact with.

REFERENCES AND FURTHER READINGS

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195). New York: Academic Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12–21.

Cattell, J. M. (1886). The influence of the intensity of the stimulus on the length of the reaction time. *Brain, 9*, 512–514.

Ehrenfels, C. von. (1890). Über Gestaltqualitäten. *Vierteljahrsschrift für Wissenschaftliche Philosophie, 14*, 249–292.

Gendler, T. S., & Hawthorne, J. (Eds.). (2006). *Perceptual experience*. New York: Oxford University Press.

Graziano, M. S. A., & Gross, C. G. (1995). The representation of extrapersonal space: A possible role for bimodal, visual-tactile neurons. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1021–1034). Cambridge, MA: MIT Press.

Leigh, R. J., & Zee, D. S. (1999). *The neurology of eye movements* (3rd ed.). New York: Oxford University Press.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Mozel, M. M., Smith, B., Smith, P., Sullivan, R., & Swender, P. (1969). Nasal chemoreception in flavor identification. *Archives of Otolaryngology, 90*, 367–373.

Myers, D. G. (2007). *Psychology* (8th ed.). New York: Worth.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53–81.

Palmer, S. F. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3*, 519–526.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 61*, 12–21.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics, 16*, 283–290.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*, 1–29.

Winkler, I., & Cowan, N. (2005). From sensory to long-term memory: Evidence from auditory memory reactivation studies. *Experimental Psychology, 52*, 3–20.

20

Psychophysics

John H. Krantz
Hanover College

Measurement is an important catalyst for the advancement of science. As measurement techniques are developed, so is the ability of scientists to ask and answer important questions. Nowhere is this observation clearer than in the development of psychophysics. Psychophysics comprises a set of methods that allow for the precise determination of how a person's internal experience relates to external stimuli. The term *psychophysics* can be broken into two parts: *psycho-*, taken from *psychology*, refers to a person's internal experience, and *physics* refers to the features of the natural world to which the person is being exposed—that is, the stimulus. The combined word is thus defined as the measurement of how a person's experience is related to changes in the physical world. Psychophysics is not just the statement that such measurement is possible; it is also an approach to this problem that makes psychophysics a fundamental advance in the field of psychology. The key to psychophysics is in how the questions are asked of the observer. It is an old saw to ask if the red you see is the same as the red some other person sees. But in a way, it is this problem of trying to access the ephemeral nature of our experience that is at the heart of psychophysics. Psychophysics does not seek to ask a person to explain his or her experience of red, but to see if different people experience red in regard to the same physical stimuli. This chapter covers the basic history and methods of psychophysics—traditional methods and the method derived from signal detection theory. Then, I discuss some of the basic psychophysical laws and illustrate the power of psychophysical methods in a few of the many applications that use psychophysics or have been built on psychophysically collected data.

A VERY BRIEF HISTORY

The founding of psychophysics dates to 1860, when Gustav Fechner published *Elements of Psychophysics* (Boring, 1950). In his book, Fechner described the basic methods for measuring psychological experience in relation to physical phenomena. In addition, he proposed that these experiences can be reduced to scientific laws.

The next major advance in psychophysical methods came with work by S. S. Stevens (1956). Psychophysics up to that time emphasized the measurement of thresholds and matches. Important as these techniques have proved to be, they lacked the ability to measure stimuli and differences between stimuli that are easily detected. Stevens, based upon a conversation in an elevator with a colleague, developed magnitude estimation that allows such measurements to be made directly.

Another important development in psychophysics comes with the advent of signal detection theory (Green & Swets, 1966). Signal detection theory is a wholly different approach to how to think about our sensitivity and detection of events in the environment. Whereas many psychologists using psychophysics will happily slide between traditional methods and signal detection theory, they actually propose fundamentally different understandings of how observers detect sensory events.

GENERAL APPROACH

Psychophysics is characterized by a general approach to collecting data. The methods tend to be very repetitive, even tedious, from the perspective of the observer. As in any psychological method, the goal is to elicit responses from the participants that are as clearly interpretable as possible. Given that so many psychological phenomena are mental experiences, this is a difficult goal. Often researchers must validate the results indirectly, much as studies of molecular phenomena must do because atoms and molecules cannot be observed directly. The approach of psychophysics is to ask questions or seek responses that require very simple decisions on the part of the observer. For example, a researcher could present a stimulus and ask the observer if he/she heard the stimulus. In other cases, the researcher might ask the observer to adjust one stimulus until it was identical to a second stimulus.

Given such simple questions and responses, it may be clear what the observer means. Still, the researcher has learned little from any single response. Thus, the experimenter needs to change the stimulus in some fashion and repeat the question(s). Usually there are many trials (questions) in the experiment. Each time, the observer is asked the same question, with the stimulus changing in systematic and often very minor ways. This feature of psychophysical experiments gives them both their power and their repetitive nature.

BASIC MEASURES

Using these procedures, researchers have measured several basic features of psychological experience.

Thresholds

A threshold, in common parlance, is a boundary. There is a threshold at a door that defines the boundary between being in the room and being outside the room. There is a direct analogy between the threshold of a room and a sensory threshold. In sensory terms, the threshold defines a boundary between perception and nonperception. Just as there is an ambiguous point as you enter a room where it is not clear whether you are in or out of the room, there is usually a small range of intensities where people will sometimes report that they perceive and at other times report that they do not perceive the same stimulus.

Absolute Threshold

The absolute threshold is the minimal level of energy that the sensory system reliably detects—for example, the dimmest light or least intense sound that can be detected. Thus, studies measuring absolute thresholds are *detection* experiments.

Difference Threshold

A difference threshold is the minimal difference between two stimuli that an observer can reliably discriminate. The ability to tell that one sound is just louder than another is a *difference threshold* or *just noticeable difference* (JND). Whereas absolute thresholds measure detection, difference thresholds measure *discrimination*.

Point of Subjective Equality

The point of subjective equality (PSE) is reached when two stimuli appear identical in some manner to the observer. For example, there is the classic illusion of length in the Müller-Lyer illusion (Figure 20.1). In this illustration, both lines are the same length, but most people say that the figure on the right with the arrowheads pointed in toward the line looks longer. To measure the PSE, the length of the line on the right could be adjusted until it appeared to the observer to be the same length as the line on the left.

BASIC PSYCHOPHYSICAL METHODS

In this section we discuss several of the most basic psychophysical methods.

Traditional Methods

The traditional psychophysical methods comprise some of the most basic and oldest methods that psychological researchers still use. Although most modern studies use more recently developed methods that avoid some of their limitations, researchers still use these methods. They provide a clear description about how psychophysics

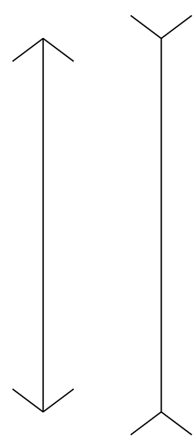

Figure 20.1 The classic Müller-Lyer illusion.

approaches the questions of measurement, which makes them valuable to cover even if they are not as commonly used as they used to be.

Method of Limits

The method of limits proceeds in a very direct fashion. The researcher preselects a set of stimulus intensities. For example, a researcher is interested in determining the finest pressure that can be felt on the forefinger (i.e., the absolute threshold for pressure on the forefinger). To begin, the researcher selects a set of intensities. A set of five to nine intensities ranging from above to below the threshold that is being measured is most common. The greatest intensity ought to be felt all of the time and the lowest intensity ought never to be felt.

The researcher then presents the stimuli to the observer in an ordered series, called a staircase. In most experiments, the researcher uses two staircases, one going from greatest to least intensity (a descending staircase) and the other going from lowest to greatest intensity (an ascending staircase). In the descending staircase, the greatest intensity stimulus is presented first and the observer is asked if the stimulus is perceived. If the experiment is set up correctly, the observer ought to be able to perceive this first stimulus. Then, the next most intense stimulus is presented and the observer is again asked if the stimulus is perceived. This procedure is repeated with each successively less intense stimulus. Eventually, a point is reached where the observer will report that the stimulus is not perceived. At this point, the next step depends upon the rule used for determining that the stimulus is no longer perceived. Two simple, common rules define the stimulus as no longer able to be perceived: (a) the first time a stimulus is reported as not perceived, or (b) the first time that two successive stimuli are reported as not perceived. The first rule, which could be called the one-stop rule, is useful in situations when the observers have very reliable responses or for rough determinations of the threshold. The second rule, which could be called the two-stop rule, is useful if finer precision is needed. With either rule, the intensity where the observer can no longer detect the stimulus is called a *turnaround*. The intensity of the turnaround is the stimulus the observer does not detect for the one-stop rule or the first of the two stimuli the observer does not detect for the two-stop rule. After this staircase is completed, the next procedure—usually an ascending staircase starting with the least intense stimulus—is begun.

The ascending staircase is just the opposite of the descending staircase; the researcher uses the same rules only now they refer to the intensities when stimuli are first detected. There are usually several sets of ascending and descending staircases to determine any one threshold. The more staircases the researcher conducts, the more precise the estimate of the threshold. The threshold is defined as the average of all of the turnarounds.

Method of Constant Stimuli

The method of constant stimuli is initially set up just like the method of limits. The researcher selects a set of five to nine stimulus intensities for the experiment. It is desirable that the most intense stimulus be easily detected and the least intense stimulus be rarely detected. The method of constant stimuli differs from the method of limits in the order of the stimulus presentation; in this method the researcher presents the set of stimuli in a random order a set number of times. Using the skin pressure example, each pressure in the stimulus set might be presented 30 times with the order completely random. Each time researchers present a stimulus, they ask the observer if the stimulus was perceived.

After running all of the trials in a method of constant stimuli experiment, the researcher will have collected the percentage of time that each stimulus in the set was detected. A sample set of data for the skin pressure example is shown in Table 20.1 and plotted in Figure 20.2. This function in Figure 20.2 is called a *psychometric function*, and it represents the change in detectability as stimulus intensity increases. However, it is still necessary to determine a threshold from this data. As you can see, there is no clean step in stimulus intensity from where the stimulus is never detected to where it is always detected. Traditionally, researchers define the stimulus that is detected 50 percent of the time as the threshold. As can be seen in Table 1, it is rare that observers actually will detect any stimulus in the set exactly 50 percent of the time. In that case, it is necessary to use some method of estimating the stimulus intensity that would be detected 50 percent of the time. There are many such methods, the simplest being to do a linear interpolation between the two intensities that were detected just below and just above 50 percent of the time. This method identifies the intensity associated with where the line between the 4th and 5th stimuli in Figure 20.2 crosses the 50 percent level on the *y*-axis.

Method of Adjustment

Method of adjustment is just as its name implies: The observer adjusts or directs the researcher to adjust the

Table 20.1 Sample data from a method of constant stimuli experiment

Pressure	*Percent Detected*
1 (Least Intense)	10
2	15
3	24
4	48
5	67
6	90
7 (Most Intense)	97

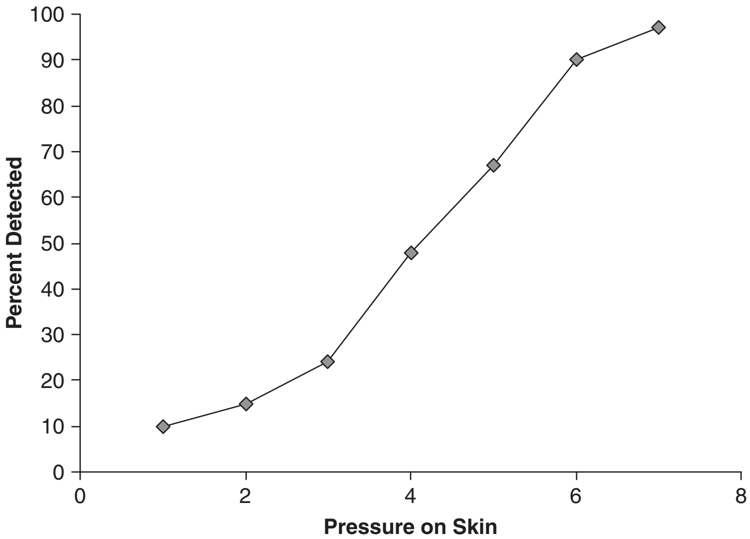

Figure 20.2 Data from Table 20.1 plotted to show how these data can represent a psychometric function of the observer's responses to all of the stimuli tested.

stimulus to the threshold. To use method of adjustment to measure a threshold, the instructions to the observer are to adjust the stimulus until it is just detectable, in the case of an absolute threshold, or just discriminably different from a comparison stimulus, in the case of a JND. However, the method of adjustment is often the method of choice to measure the point of subjective equality. Whereas researchers can use the method of limits and method of constant stimuli to infer the point of subjective equality, with the method of adjustment, the researcher can ask the observer to directly match two stimuli. To take the Müller-Lyer illusion example in Figure 20.1, the researcher can instruct the observer to adjust the length of the line on the right until it matches the length of the line of the left.

Whether measuring a threshold or PSE, the observer repeats the adjustment several times. It is important to have the starting value of the stimulus to be adjusted—for example, the length of the line on the right in the Müller-Lyer example. In this way, the observer must make a fresh adjustment based upon his or her perception instead of making adjustments based on the memory of previous trials. The threshold or PSE is the average of the final settings from each trial.

Limitations of Traditional Psychophysical Methods

The primary criticism of these methods is that they are dependent on the report of the observer. There is no direct way to assess whether the observer's report is accurate or not. It is possible that observers may seek to make their sensory abilities seem greater than they are and will sometimes report that they detect a stimulus even when they do not. Other observers might be cautious and want to be completely sure that a stimulus is present before they report that they detect it.

The problem remains that these methods rely on the accuracy of the response of the observer. One correction was to put in *blank* trials. In an absolute threshold experiment, a blank trial would be a trial without a stimulus; in a difference threshold experiment, a blank trial would include two stimuli that are identical. From the response to these blank trials, the results from the actual trials could be adjusted. However, as you will see, researchers sought better ways to deal with this issue. Forced-choice methods and signal detection theory, discussed below, are different ways of dealing with this problem.

Another limitation is that these methods measure only thresholds or PSEs; they cannot allow direct measurement of how perceptible a stimulus is that is easily perceived or how different two stimuli are that are easily discriminated. Magnitude estimation provides a means to make these measurements.

Newer Methods

Forced-Choice Method

In order to obtain a clearer measure of the observer's sensitivity to a stimulus, the forced-choice method changes the way the researcher structures a trial and the question asked of the observer. I will illustrate a forced-choice method with two alternatives. Using the traditional psychophysical methods to measure the absolute threshold, the researcher presents the stimulus alone and asks the observer if he or she detected the stimulus. The researcher using a forced-choice method has two options: a *temporal forced-choice trial* or a *spatial forced-choice trial*. In a temporal forced-choice trial, there are two time periods—Time Period A is shortly followed by Time Period B. The researcher randomly presents the stimulus during one of the two time periods and asks the observer during which time period the stimulus occurred. If the observer is unsure, then he or she must guess; this is the forced feature of the forced-choice method. In a spatial forced-choice trial, there are two locations—for example, the left or right ear in an auditory study. The observer must indicate in which ear the stimulus is presented.

For the rest of the experiment, the forced-choice method can be combined with either the method of limits or the method of constant stimuli. If the forced-choice method is combined with the method of limits, researchers define the turnaround as the moment when the observer guesses incorrectly when or where the stimulus is presented. In the method of constant stimuli, the threshold has to be defined differently. Guessing correctly 50 percent of the time in a forced-choice method is what happens when the observer cannot detect the stimulus at all in a two alternative, forced-choice situation. So, when researchers use a two alternative, forced-choice method combined with the method of constant stimuli, they adopt 75 percent as the value for the threshold.

Magnitude Estimation

The essence of measurement is the assigning of a number to a quantity. For the measurement to be valid, the assignment of the number must have a meaningful relation to the phenomenon under investigation. Temperature assigns a number related to the amount of heat in a system, and the values have proved meaningful. Generally, finding a useful measure is a difficult task. Every now and then, the simplest procedure works to provide useful measures of phenomena of interest. Magnitude estimation, developed by S. S. Stevens (1956), is one such simple method. The psychophysical methods discussed previously have specified elaborate means for determining a measure such as a threshold. In magnitude estimation, the researcher presents a stimulus of a given intensity and observers assign a number that represents how strongly they experience the stimulus. For example, in a measure of perceived intensity of a sound stimulus, a sound is presented and the observer replies with a number that represents how loud that sound is. It does not matter that different observers might use different number ranges; researchers use basic range correction procedures, akin to changing from Fahrenheit to Celsius, to put the observers' magnitude estimates all in the same range.

However, if the researcher desires to restrict observers' estimates to a single range of values, it is possible to take one stimulus intensity and make it a standard intensity. This intensity is presented to the observer and given a value, say 100 or 50. This standard intensity is called a *modulus*. The use of the modulus tends to restrict observers to using more similar ranges of estimates, though researchers often still range correct these estimates (e.g., Silverstein, Krantz, Gomer, Yeh, & Monty, 1990).

Despite the simplicity of this approach, it yields reliable and remarkably consistent results (Krantz, Ballard, & Scher, 1997; Silverstein et al., 1990). Subsequently I will discuss how this method has changed ideas on the perception of stimuli of different magnitudes. Reflecting one of the hallmarks of a good method, magnitude estimation has changed ideas and led to a more flexible understanding of how senses handle stimuli of different magnitudes.

Signal Detection Theory

Signal detection theory, like forced-choice methods, is interested in measuring the sensitivity of the observer to the stimulus separate from any response bias, but it takes a very different approach. In essence, the discussion of thresholds has assumed that the detection of a stimulus, at its basic level, is merely an automatic response to the stimulus. However, anyone who has been in a situation where they have tried to detect a weak stimulus knows that there is nothing automatic about detecting such a stimulus; consider listening for a strange sound a car motor might be making while traveling down the road. It is ambiguous whether that sound is there or not; different people in the car might disagree on the presence of the sound. Even if everyone agrees that the sound is there, it is still possible to disagree about what type of sound it might be and what it might portend about the condition of the car. Detecting a stimulus is not automatic and involves a person's cognitive abilities.

Signal detection theory agrees with this intuition and describes stimulus detection as a cognitive event. In signal detection theory, the stimulus is called a *signal*. Although this change in terminology is a bit confusing, it recognizes the history of signal detection theory, which comes to psychology via engineering. In this field, the to-be-detected event often was a phone signal transmitted over phone lines; this terminology carried over to psychology.

Signal Detection Theory Experimental Description and Trial Outcomes

The signal is not presented on every trial. On each trial, the observer's task is to report whether the stimulus was presented or not. Because the stimulus is not presented all of the time, the observer might or might not be correct. Table 20.2 presents the possible results of any trial in a signal detection theory experiment. First, consider the trials in which the researcher presents the signal. The outcomes of these trials are shown in the left-hand column of Table 20.2. If the signal is presented, the observer may report that the signal was presented. In this case, the observer is correct, and this type of trial is called a *hit*. Even though the signal is presented, the observer may report that the signal is not presented. This situation is represented in the lower left cell of Table 20.2. The observer's response is incorrect, and the trial is referred to as a *miss*. Next, consider the trials in which the signal is not presented. The outcomes of these trials are shown in the right-hand column of Table 20.2. When the signal is not presented, but the observer responds that it was, the trial is called a *false alarm*. This is the boy-who-cried-wolf type of trial. It is also an error like the miss trials. If the signal is not presented and the observer responds that the signal has not been presented, then the trial is called a *correct rejection*, as shown in the lower right cell of Table 20.2.

For theoretical reasons, the responses to trials in which the signal is presented and the responses to trials in which the signal is not presented are tracked separately. Thus, the percentages of trials that are hits and the percentage of trials that are misses add up to 100 percent of trials in which the signal is presented. In the same way, the percentage of trials that are false alarms and the percentage of trials that are correct rejections add up to 100 percent of trials in which the signal is not presented. One practical

Table 20.2 Stimulus presentation, participant response, and outcomes in signal detection theory

	Is the stimulus presented	
Participant response to question: Did you perceive the stimulus	*Yes*	*No*
Yes	Hit	False Alarm
No	Miss	Correct Rejection

advantage of this method of counting trials is that it is possible to just track hits and false alarms and know how the observer responded to misses and correct rejections. Similarly, it creates fewer numbers that have to be tracked by the researcher.

Underlying Theoretical Explanation for Outcomes in Signal Detection Theory

Signal detection theory begins its explanation for the way people will react in these experiments by arguing that in any perceptual system there is some level of activity, even in the absence of a signal. Sometimes this activity is greater; other times it is lesser; it is not constant. According to signal detection theory, the activity in the perceptual system that occurs when there is no signal is called *noise* (see Figure 20.3) and varies around a mean level according to a normal distribution. Noise is always present and never goes away, even when the signal is presented.

The signal simply adds a fixed amount of activity to the current level of noise in the perceptual system. For example, if there is no noise in the perceptual system when the signal is being processed, the perceptual system will respond with a level of activity in exactly the same amount as the effect of the signal. If there is a large amount of noise in the perceptual system when the signal is being processed, then the perceptual system will respond with a level of activity equal to the sum of the noise and the effect of the signal. The result is that the effect of the signal on the perceptual system is not fixed but varies exactly to the same degree as the noise, but with a mean level of activity defined by the sum of the mean activity of the noise and the effect of the signal. Researchers call the curve in Figure 20.3 that represents what happens when the stimulus is presented the *signal + noise curve*.

The distance between the noise curve and the signal + noise curve indicates the sensitivity of the observer to the stimulus. When this distance is measured in the number of standard deviations between the signal + noise curve and the noise curve, this measure is called *d'* (Figure 20.3). When the observer's sensitivity to the signal is not great, the noise curve and the signal + noise curve overlap, as shown in Figure 20.3, and there are some levels of activity in the perceptual system that could be caused by noise alone or by the presence of the signal. Only one of these events can happen at a time. The noise alone, as you will recall, happens on a trial when the signal is not presented and the signal plus noise happens when the signal has been presented. The observer in this experiment does not know which trial has been presented. Still, there are levels of activities that could come from either type of trial. In this circumstance, the observer must decide which type of trial occurred and it cannot be based on any completely objective evidence in the sensory system.

This necessity of making a decision leads to the last feature of signal detection theory, the *criterion*. The criterion represents a cutoff in the decision making. If the level of activity in the perceptual system is below the criterion, the observer will report that the signal was not presented, whether it was or not. Conversely, if the level of activity of the perceptual system is above the criterion then the observer will report that the signal was presented, again whether it was or not. One possible position for the criterion is shown in Figure 20.3. From this diagram, the four possible outcomes of experimental trials just discussed can be observed.

1. If the signal is not presented and the level of activity in the perceptual system from the noise is below the criterion, the observer will report that the signal is not present. This is a correct rejection.
2. If the signal is not presented and the level of activity in the perceptual system from the noise is above the criterion, the observer will report that the signal is present. This is a false alarm.
3. If the signal is presented and the level of activity in the perceptual system from the signal plus noise is below the

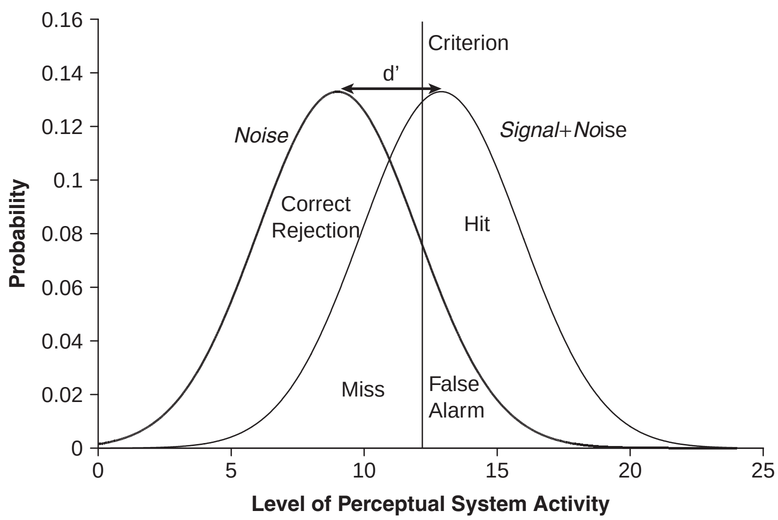

Figure 20.3 The theoretical underpinnings of signal detection theory.

criterion, the observer will report that the signal is not present. This is a miss.

4. If the signal is presented and the level of activity in the perceptual system from the signal plus noise is above the criterion, the observer will report that the signal is present. This is a hit.

One important feature of the criterion is that it is set by the person making the observations and can be set at any desired level of activity in the perceptual system. For example, if the observer really wants to maximize the number of hits, the criterion can be set to the left of its position in Figure 20.3. This change will increase the probability that the perceptual system activity will be above the criterion when the signal is present. However, it also will increase the chance for the perceptual system activity from a trial when no signal is present to be above the criterion. As a result, false alarms will also increase. This is called a *lax criterion*. Conversely, it might be that the observer wishes to minimize false alarms by moving the criterion to the right in Figure 20.3. The rate of hits would go down as well. This is called a *strict criterion*. According to signal detection theory, then, the proportion of hits and false alarms will depend upon the criterion and sensitivity. One measure of criterion is called *beta* and is the ratio of the height of the signal + noise curve over the noise curve at the criterion.

PSYCHOPHYSICAL LAWS

Psychophysical methods have lead to psychophysical laws, summaries of findings from psychophysical studies. We will examine three important laws: Weber's law, Fechner's law, and Stevens's law.

Weber's Law

Weber's (pronounced Vāber) law is the oldest and simplest of the laws. Weber's law deals with the description of the size of a difference threshold relative to the background intensity. Weber's law is given in Equation 20.1 below.

$$\text{Equation 20.1} \qquad k = \Delta I/I$$

In Equation 20.1, I refers to the intensity of a stimulus. The I alone refers to the intensity of the stimulus against which the change in stimulus intensity is being measured, sometimes called the *background intensity*. The ΔI refers to the change in the intensity that reaches the threshold. This measure is the *difference threshold* or *JND*. The k is a constant. Translating into English, Equation 20.1 states that the ratio of the difference threshold to the background intensity is a constant. The implication of this law is that observers lose sensitivity to changes in intensity as the background becomes more intense. Not only does sensitivity decrease, but it decreases at a rate proportional to the increase of the background intensity.

Although Weber's law does not hold for all situations (e.g., very weak and very intense background intensities), it does seem to hold for most situations. Moreover, this law has many important implications. Most people do have an intuitive sense that Weber's law holds, at least in a general manner. If one candle is lit in a dark room it is easy to perceive the change. But if one additional candle is lit in a room with 500 candles already lit, it is unlikely that anyone would notice the change even though the change in light intensity (ΔI) is the same in both cases. However, the ΔI/I in the case of the dark room is very large and very small in the case of the room with 500 candles lit.

Fechner's Law

Recall that the classical psychophysical methods are mainly useful for measuring thresholds, either absolute or difference. Fechner was interested not only in our threshold perception, where observers are at their limits of ability, but also in the more common suprathreshold perception, where detection or discrimination is easy. Fechner's law is given in Equation 20.2.

$$\text{Equation 20.2} \qquad S = k \times \log(I)$$

In this equation, I stands for the intensity of the stimulus, S stands for the strength of the sensation as experienced by the observer, and k is a constant. In fact, k is the constant from Weber's law for the given sensory system (e.g., the experience of light intensity). Fechner's law, by using logarithms, follows along with Weber's law that we lose sensitivity to all stimuli as their intensity grows. Fechner simply applied Weber's law about thresholds to all levels of intensities, subthreshold, near threshold, and much greater than threshold. However, there was no direct evidence to support this law, as the psychophysical methods of the day did not directly measure the magnitude of experience of suprathreshold stimuli.

Stevens's Law

Data from magnitude estimation experiments gave S. S. Stevens direct measures of observers' suprathreshold experiences. In several cases, Fechner's law seemed to hold reasonably well, as when observers determined how intensely a light or a sound was experienced. In other cases that Stevens and others tested, however, Fechner's law did not seem to hold too well, as when observers estimated line length. The relation seems to be more linear than logarithmic. What that means is that if one line is twice as long as another line, it looks twice as long. If perceptions of line length were logarithmic, the longer line would look less than twice as long. In addition, there were other examples in which sensory experience increased even faster than physical intensity. For example, when using low levels of electrical shock (at levels that do not cause harm), if the experimenter doubles the intensity of the stimulus it more than doubles the experienced pain.

Although Fechner's law seemed adequate in some cases, it is always preferable to seek a more general description that applies to all of the situations described. Stevens found one law that applied to all of the situations so far described (see Equation 20.3).

Equation 20.3 $S = cI^b$

This is an exponential function. The symbols S and I are the same as in Fechner's law: strength of sensory experience and intensity of the stimulus, respectively. The symbol c is a constant but has no special meaning. The important symbol that describes the nature of the relation between S and I is b. The exponent b describes the shape of the relation between the physical stimulus and the sensory experience (see Figure 20.4). If b is less than 1, sensory experience increases more slowly than physical intensity, similar to Fechner's law. In Figure 20.4, which has an example of Fechner's law on it as a reference, the $b < 1$ curve becomes ever flatter, without ever reaching an asymptote. The shape is not identical to that for Fechner's law, though it can be made more similar than is indicated on Figure 20.4, but the similarity of the two types of relations is clear. The perception of light intensity and sound intensity follows this type of relation, which is why Fechner's law worked reasonably well.

When $b = 1$, Equation 20.3 becomes the equation of a line as shown in Figure 20.4. Thus, the perception of line lengths would follow this form of Stevens's law. For $b > 1$, the form of Stevens's law is the curve that has an ever-increasing slope to it. The perception of pain to electric shock follows this pattern. Given the observation that Stevens's law can be used to describe a wider range of results, it is generally the preferred law for describing these relations.

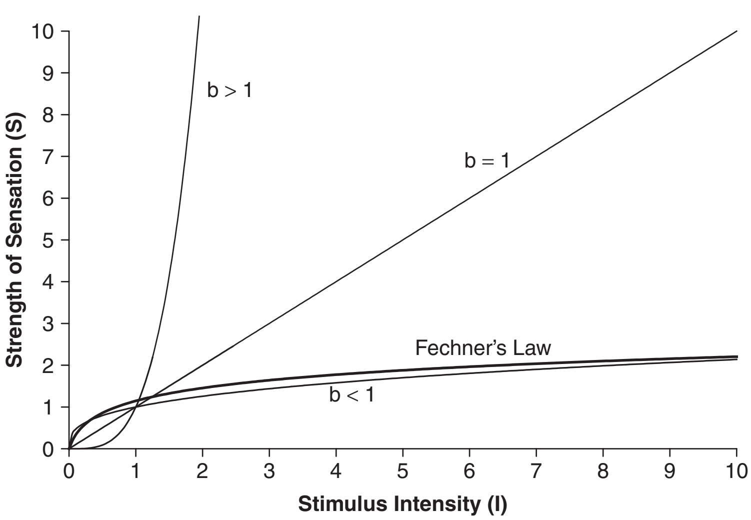

Figure 20.4 A comparison of the possible relationships between stimulus intensity (I) and sensory experience (S) in both Stevens's law and Fechner's law.

APPLICATIONS

Since their introduction, researchers have applied psychophysical methods to countless practical problems; as a result, many technologies have been impacted.

Development of the Color Matching System (CIE)

One of the major technological developments of the 20th century was the development of color reproduction technologies. The ability to reproduce color in film and print seems standard and everyday, but the development of this technology was a concerted effort during the early part of the 20th century. Although many disciplines play very important roles in this development, at the center is the need to clearly and precisely describe how color mixing and matching happen in the human visual system. Understanding how the human visual system deals with color mixing is vital because it forms the basis of color reproduction technologies. Rather than re-create the exact physical complement of wavelengths in the original object, these technologies produce a set of wavelengths that causes the human observer to perceive the same color (see Chapter 23). Researchers used psychophysical methods, especially the method of adjustment, to define how color mixing occurs in what was defined as the standard color observer. This standard color observer was turned into a mathematical system that allows color matches to be made without the need for a human observer. The first such system was brought out in 1931 by the *Commission Internationale de l'Eclairage* (CIE) in France, and has been updated since (Silverstein & Merrifield, 1985). This system led quickly to practical applications such as color movies, which were first produced later in the same decade.

Development of Seat Belt Alarms

As cars have become faster and driving more dangerous, the safety of the vehicle has achieved paramount importance. Although new devices are developed every year, seat belts are still the principal safety device in the car, and finding ways to encourage their use has been a concern for many years. One recent method to try to get people to use the seat belt is a high-pitched dinging sound that indicates that the driver is not buckled. The dinging sound is relatively quiet and not too unpleasant to hear. This sound contrasts with the sound some car manufacturers use as a seat belt alarm, which is a very loud, unpleasant noise. The loud noise seemingly makes sense because it is desired that the sound never be covered up by any background noises.

However, psychophysical experiments on how noises mask other sounds revealed that this loud noise was unnecessary and only served to irritate drivers (Zwicker, Flottorp, & Stevens, 1957). Psychophysical experiments have revealed that noises can mask sounds only in the same frequency range, a concept known as a *critical band*. Background sounds made by car engines tend to be low-frequency sounds. Our voices and any music we might play tend to be of a middle range of frequencies. So by using a high-frequency but relatively low-intensity tone, the sound is unlikely to be covered up by the other noises in the car. For this reason, car manufacturers could abandon the loud, unpleasant sound for one that is far more tolerable.

Cockpit Displays

In the late 1970s and early 1980s, there was a major revolution in airplane instrumentation. Starting with the Boeing 757/767 and the McDonnell-Douglass DC9-80, all of the dials that pilots had been reading were replaced with electronic displays. In these first instances, the displays were CRTs—the same type of device as the standard, deep-back television. There were many engineering and safety goals in putting these devices into the airplane, including reliability and flexibility in what can be displayed, such as radar information. However, there were many barriers of a perceptual nature that had to be overcome to allow the use of these devices.

The main difficulties with these devices are twofold and interrelated. First, CRTs are not nearly as bright as the sun, which means that when flying into the sun, the pilot's adaptation state is so great, reducing the pilot's visual sensitivity, that the CRT can be very hard to see. Second, the surface of the CRT reflects a fair amount of light. Because the most powerful light source in the sky is the sun, it tends to wash out everything on the screen. All of the colors become more desaturated (see Chapter 23) and more like white; hence, it is very difficult to tell one color from another. This fact can be quite a problem because the colors mean different things to the pilot. For example, red, amber, and cyan each indicate different levels of problems with the plane (red is for the most severe problems; Silverstein & Merrifield, 1985).

A series of psychophysical experiments conducted by Silverstein, among others, identified the intensity needed by the CRT in all conditions to allow the pilot always to be able to read the screen. In addition, these studies identified a set of six colors that could be distinguished even under the worst conditions. Finally, this psychophysical data allowed engineers to devise an automated mechanism to adjust the intensity of the CRT so that it was at an appropriate light level for all different types of lighting conditions so that the pilot did not have to keep adjusting the display intensity (Silverstein & Merrifield, 1985). This work was recognized with a national award by the Human Factors Society.

Quick Sensory Screening

Psychophysical methods tend to be very time-consuming and are unsuited for health-related screening in normal practice. For example, many states require school systems to screen students for vision and hearing problems. The number of students who have to be tested in a timely fashion requires a faster method of testing than the traditional psychophysical methods allow. So these methods need to be adjusted to allow for a more rapid determination of results. With simplifying the methods, some precision in the results will be lost. For many of these purposes, the loss of precision is acceptable. Here are two examples.

Vision Screening and Getting Glasses

Most people are familiar with having their vision tested; it is required whenever you apply for a driver's license. These tests screen acuity, the ability to resolve fine details. In particular for driving, the acuity for objects at a distance is tested. Many of these tests are based on a method of screening devised by Dutch ophthalmologist Hermann Snellen in 1843. The chart has letters with carefully calibrated features that, when they are too small, cause the letter to become confused with other letters. The observer reads the chart, usually starting with large letters, and continues until the letters start to become confused. This screening is a variation of a method of limits with a descending staircase. At each level there is some repetition of a stimulus of the same size to gain some greater precision.

Auditory Screening

Auditory screenings, when needed to be done quickly, often use either a variation of a method of limits or a forced-choice method. In the method of limits approach, the tone is decreased in intensity until the child no longer hears the sound. Sometimes the tone starts from a very low intensity and the intensity is increased until the child can hear the sound. Often this staircase is not repeated and the next frequency is tested directly. In this way, a profile of the sensitivity across a range of frequencies can be determined in a relatively short time. Although subtle differences in auditory sensitivity may not be detected, large-scale problems that could affect a student's performance in school can be discovered.

Sometimes the requirement is only to test a single critical intensity at several frequencies. In this case, a modified forced-choice method is desired. The student, wearing headphones, will be asked in which ear he or she hears the sound. By repeating the trial a few times, it is possible to determine quickly if this tone can be heard.

FUTURE DIRECTIONS

Psychophysical methods are not a static set of ways to collect data. As in all of science, the methods used are being adapted and modified constantly. In one direction, methods are being adapted for use in measuring a wider range of perceptual phenomena than the threshold. In another direction, these methods are being adapted for use in wider areas of psychology and are beginning to fulfill Fechner's ideas when he first proposed this methodology.

Thresholds are important but they represent the limits of sensation and perception. We need to learn more about the sensation and perception that occur during the more normal operation of perceptual systems. Object recognition is an area of intense current research interest—for example, what are the perceptual features of a stimulus that allow it to be recognized as a dog? One example from this research area will illustrate this point. Collin, Therrien, Martin, and Rainville (2006) were interested in how blurring (filtering) images would play a role in the perceptual information needed to correctly identify a face. The filtering of the faces was carefully designed to match some theoretical proposals about how the visual system processes faces. For the current purposes, the method is more important. In this experiment, one face was presented that was filtered. There were four other faces, one of which was the same as the filtered face. The observers in this experiment were to adjust the filtering of the original face until they could correctly identify which of the other four faces it was. The adjustment of the filtering uses the method of adjustment; however, neither a threshold nor a point of subjective equality is really being used. Here, the observer is asked to identify a stimulus in a modification of a forced-choice situation using four alternatives.

Psychophysical methods are also being applied to a wider range of psychological areas and are not as exclusively a method of sensation and perception. For example, cognitive and social psychologies have found use for psychophysical methods. Keeping in mind the topic of facial recognition, it is clear that this issue also is relevant to eyewitness identification. In one study, modifications to the face such as adding eyeglasses or removing a beard reduced accuracy of identification. This reduction in accuracy was measured using d' from signal detection theory (Terry, 1994).

SUMMARY

This chapter has examined the nature of psychophysical methods. These methods have allowed for the precise measurement of sensory and other psychological phenomena. These methods require careful manipulation of the stimuli with many repetitions of the trial. Despite their longevity, the classical methods still find application (Collin et al., 2006). Signal detection theory represents the greatest departure in approach in psychophysics since its inception. Signal detection theory sees even the basic detection of a sensory stimulus as a cognitive event requiring decisions on the part of the observer. The precision of the psychophysical methods has allowed for application of knowledge about how sensory systems work to be used in many diverse areas. As all methods do, psychophysical methods will continue to develop and be applied to an ever-wider range of sensory phenomena.

REFERENCES AND FURTHER READINGS

Boring, E. (1950). *A history of experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.

Collin, C. A., Therrien, M., Martin, C., & Rainville, S. (2006). Spatial frequency thresholds for face recognition when comparison faces are filtered and unfiltered. *Perception & Psychophysics*, *68*, 879–889.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, *29*, 264–269.

Silverstein, L. D., Krantz, J. H., Gomer, F. E., Yeh, Y., & Monty, R. W. (1990). The effects of spatial sampling and luminance quantization on the image quality of color matrix displays. *Journal of the Optical Society of America, Part A*, *7*, 1955–1968.

Silverstein, L. D., & Merrifield, R. M. (1985). The development and evaluation of color systems for airborne applications (DOT/FAA/PM-85-19). Washington, DC: DOT/FAA.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, *69*, 1–25.

Terry, R. L. (1994). Effects of face transformation on accuracy of recognition. *Journal of Social Psychology*, *134*(4), 483–492.

Zwicker, E., Flottorp, G., & Stevens, S. S. (1957). Critical bandwidth in loudness summation. *Journal of the Acoustical Society of America*, *29*, 548–557.

21

States of Consciousness

Joseph J. Palladino and Christopher M. Bloom
University of Southern Indiana

The term *consciousness* has a long history in psychology. One way to assess the concept is to consider how it has been addressed in our discipline's gateway—introductory psychology texts. The first edition of the classic R. C. Atkinson and R. L. Atkinson textbook appeared in 1957. The tenth edition of this text was published in 1990. The topic of consciousness occupies a chapter in a section of the text titled *Consciousness and Perception*; the specific chapter that deals with the topic is titled *Consciousness and Its Altered States*. The breadth of topics in this chapter is similar to what we find in more recently published textbooks (e.g., Davis & Palladino, 2007; D. H. Hockenbury & S. E. Hockenbury, 2006) and includes subtopics such as aspects of consciousness (conscious, unconscious, subconscious), divided consciousness, dissociation and multiple personality, sleep and dreams, psychoactive drugs, meditation, and hypnosis. The authors begin their chapter with these words:

> As you read these words, are you awake or dreaming? Hardly anyone is confused by this question. We all know the difference between an ordinary state of wakefulness and the experience of dreaming. We also recognize other *states of consciousness*, including those induced by drugs such as alcohol and marijuana. (R. L. Atkinson, R. C. Atkinson, Smith, & Benn, 1990, p. 195)

Despite the apparent clarity of the opening of the Atkinson et al. (1990) chapter on consciousness, the topic is perhaps one of the more difficult ones for the discipline, as its history reveals. Wilhelm Wundt's (1832–1920) founding of the first psychological laboratory in 1879 heralded his attempt to identify and categorize elements of conscious experience using the method known as *introspection*, or examination of one's own mental state. Wundt's student, Edward Bradford Titchener (1867–1927), brought Wundt's approach to the United States and initiated a perspective known as *structuralism*. Another early psychologist, William James (1842–1910) also expressed interest in consciousness. He wrote about the *stream of consciousness* and described it as continuous, changing, and having depth. In contrast to Wundt and Titchener, James did not focus on analyzing and reducing conscious experience to its supposed elements. To him, attempting to divide this stream would distort the unity of conscious experience.

Despite the close connection between consciousness and the founding of the discipline, the topic of consciousness would fall into disfavor for quite some time. The development of the behavioral movement, led by John B. Watson (1878–1958), focused on observable behaviors; there was no allowance for the nonscientific discussion of what could not be observed—conscious experience. In 1913 Watson wrote, "Psychology as the behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its methods" (p. 158). Thus, the study of consciousness was, if not banned, certainly placed firmly on the back burner for several decades.

Nevertheless, advances in techniques and monitoring equipment have made it possible to conduct more sophisticated research than Wundt or Titchener would have ever thought possible. For example, in 1929, Hans Berger developed the *electroencephalograph* (EEG), which monitors and records brain activity through electrodes attached painlessly to the skull. The common belief that the EEG provides electrical stimulation to the brain is a myth. The brain's electrical signals are amplified and printed out, producing a record of brain waves. Several commonly observed brain waves are *alpha*, *beta*, *theta*, and *delta* (more on brain waves later in this chapter).

Although the EEG was a major advance, it does have several limitations. For example, although brain waves provide some evidence of activity occurring in the brain, this information is generally an imprecise picture of brain activity. One could compare the information obtained from an EEG to what we know when we pass outside a football stadium and hear the crowd noise as they react to the game. You might have a general notion of what occurred but you would find it difficult to provide specifics concerning the activity that is occurring inside that stadium.

Nonetheless, the EEG does have its uses—for example, it is helpful in diagnosing epilepsy, in identifying certain sleep disorders such as narcolepsy (National Institute of Neurological Disorders and Stroke, 2006; Honma et al., 2000), and in sometimes providing information on the presence and location of brain tumors. An EEG also can be used to confirm a state called *brain death*, in which a flatline EEG tracing indicates the absence of brain activity (Afifi & Bergman, 2005). The more recently developed and highly sophisticated *magnetoencephalography* (MEG) measures the brain's magnetic fields and can determine levels of electrical activity in a more precise manner than a standard EEG (Huettel, Song, & McCarthy, 2004).

The development of computers has been responsible, in part, for a major leap forward in the understanding of brain activity during various states of consciousness. Scientists interested in consciousness no longer have to rely on measures such as the EEG. In fact, "Where questionnaires, interviews, and observations of behavior once reigned supreme, fancy machines now create portraits of brains at work" (Bower, 2002, p. 251). Newly developed techniques can even record the activity of a single neuron; however, in general practice they produce various images of the brain (some static and some of ongoing activity). These brain-imaging techniques are significant advances for diagnosis and research, especially in comparison to the EEG and the X-rays of the brain that have been used in the past. Here is a list of some of these advanced brain-imaging techniques:

Positron emission tomography (PET): a dynamic scan that can provide evidence of ongoing brain activity by measuring the brain's metabolic activity.

Computerized axial tomography (CT or CAT): a brain-imaging technique that involves computer interpretation of a large number of X-rays.

Magnetic resonance imaging (MRI): uses a strong magnetic field and radio waves to produce very detailed pictures of the brain. Functional magnetic resonance imaging (fMRI) is a modified version of the MRI that is capable of providing both excellent structural views of the brain and ongoing changes in brain activity. (Huettel et al., 2004)

WHAT IS CONSCIOUSNESS?

A common definition of consciousness is "personal awareness of feelings, sensations, and thoughts at a given moment" (Davis & Palladino, 2007, p. 133). Our consciousness, or awareness, can be said to vary across a continuum ranging from complete awareness at one end of the spectrum to coma, vegetative states, and brain death or complete lack of awareness at the other end. Throughout the day, we experience changes in our consciousness, from full involvement and engagement in a task to states of drifting consciousness (How many of us have no memory of having driven ten miles since the last exit?) and finally to sleep or perhaps other loss of consciousness such as during surgery, a coma, or a vegetative state.

Various forms of fantasy and imagination are common examples of changes in our level of awareness. The change in consciousness we call daydreaming is one such common experience. The frequency of daydreaming is highest among children and tends to drop with age (Giambra, 2000). In children, fantasy life may extend to the creation of imaginary companions, which is quite common (Bower, 2005).

Who have not found themselves daydreaming while driving or listening to a lecture? Most daydreams are spontaneous images or thoughts that pop into our mind for a brief time and are quickly forgotten. Some people can use their daydreams to solve problems or to rehearse a sequence of future events. Nevertheless, most daydreams are related to everyday events such as deciding what we will eat for dinner later in the day or thinking about the sports car we would love to drive if only we could afford it! In fact, about two thirds of daydreams are connected to our immediate situation and surroundings. Contrary to popular belief, only a small portion of daydreams involve sexual content (Klinger, 1990), although approximately 95 percent of men and women report having had sexual daydreams at some time (Leitenberg & Henning, 1995).

Daydreams would seem to be rather elusive phenomena to study because they are hard to predict and hard to capture. Consequently, researchers have had to be creative in devising methods to study daydreams. One especially useful technique involves equipping people with beepers designed to sound at random intervals to signal them to report their daydreams in written form (Klinger, 1990). Researchers have also used physiological measuring devices to study daydreams; daydreaming seems to be associated with changes in the ratio of different brain waves (Cunningham, Scerbo, & Freeman, 2000).

Hypnosis

One of the most controversial topics in psychology may be hypnosis. The connection to stage hypnosis often makes the public wonder if there is anything behind what is often perceived as just stage effects designed for entertainment purposes (Streeter, 2004). We can trace the history of hypnosis to Franz Anton Mesmer (1734–1815), an Austrian physician, who captured the imagination of residents of Paris by claiming that he could cure anything that was ailing them. He believed the atmosphere was filled with invisible magnetic forces he could harness for their curative powers; people with a variety of ailments sought his supposed healing powers. His efforts to harness those supposed magnetic forces became known as a technique called *mesmerism*. Despite numerous testimonials in support of his treatment efforts, in 1784 a scientific commission chaired by Benjamin Franklin determined that what Mesmer seemed to achieve was the result of his patients' imagination, not invisible magnetic forces. Nevertheless, his techniques survived in the hands of James Braid (1795–1860) a Scottish surgeon who changed the name from mesmerism to *hypnosis*. Although the word is derived from the name of the Greek god of sleep, hypnosis is not sleep. The French physician Jean-Martin Charcot (1825–1904) studied hypnosis, and as a result it developed a degree of respectability as an area of medicine. Sigmund Freud learned to use hypnosis from Charcot, and used it as a treatment technique early in his career.

A common definition of hypnosis is "a social interaction in which one person, designated the subject, responds to suggestions offered by another person, designated the hypnotist, for experiences involving alterations in perception, memory, and voluntary action" (Kihlstrom, 1985, p. 385). Historically, researchers have used the concept of suggestibility or susceptibility to explain phenomena subsumed under the term hypnosis. The term reflects the degree to which a person follows suggestions offered by the hypnotist. In a typical hypnotic treatment, the hypnotist creates a situation in which the hypnotized person is more likely to follow his or her suggestions. The process of putting someone in a hypnotic state, called a *hypnotic induction*, usually involves having the person stare at an object (such as a watch), inducing relaxation, and encouraging drowsiness. A person's degree of suggestibility to hypnosis can be assessed by using measures such as the Stanford Hypnotic Susceptibility Scale, which consists of a series of 12 activities designed to assess the depth of the hypnotic state. For example, the person may be told that he or she has no sense of smell, then the hypnotist will wave a vial of ammonia under the person's nose. The person's subsequent reaction to the ammonia reveals his or her responsiveness to hypnosis (Nash, 2001). Scores on measures of suggestibility or susceptibility follow a normal curve (Patterson, 2004) and tend to be quite stable across time (Nash, 2001). What's more, "Hypnotizability is unrelated to personality characteristics such as gullibility, hysteria, psychopathology, trust, aggressiveness, submissiveness, imagination, or social compliance. The trait has, however, been linked tantalizingly with an individual's ability to become absorbed in activities such as reading, listening to music, or daydreams" (p. 49).

There are numerous claims surrounding hypnosis, including its ability to reduce pain, to treat addictions such as smoking and alcoholism, to overcome shyness, and to treat insomnia (Streeter, 2004). A review of 18 published studies of hypnotic pain relief indicated that 75 percent of the participants obtained substantial pain relief from hypnotic techniques (Montgomery, DuHamel, & Redd, 2000). Nevertheless, the mechanisms responsible for the reported pain reduction associated with hypnosis are not clear (Nash, 2001).

One of the most controversial aspects of hypnosis is the claimed ability to improve memory, especially for criminal activity or cases of abuse. Despite claims for its effectiveness, evidence indicates that hypnosis can lead to a situation in which false memories can be created, whether intentionally or unintentionally (Yapko, 1994). As a result, most hypnotically elicited testimony is now excluded from our court system (Newman & J. W. Thompson, 2001).

The controversy concerning hypnosis lies in attempts to explain it. One argument notes that what we label hypnotic behavior is simply a person following what he or she believes is the role of a hypnotized person. Thus, this *cognitive-social* explanation argues that hypnosis does not involve an alteration in consciousness; it is simply acting out a role. As Ernest Hilgard (1991) noted, "I would be more comfortable for the investigator if there were some precise indicator of the establishment of a hypnotic condition" (p. 30). An alternative explanation, offered by E. R. Hilgard (1904–2001), posits a process of *dissociation* in which there is a splitting of conscious awareness, which may well resemble the splitting of consciousness that occurs during everyday activities as simple as driving a car while one's mind seem to wander. More recently, researchers using PET scans (Kosslyn, W. L. Thompson, Costantini-Ferrando, Alpert, & Spiegel, 2000) asked hypnotized and nonhypnotized individuals to imagine that brightly colored shapes were actually gray; they were also shown gray shapes and asked to imagine that they were brightly colored. Under hypnosis, there were observed changes in both hemispheres rather than a change only in the right hemisphere, which suggests that there are actual brain changes that occur during hypnotic inductions. In other words, the subjective experience that occurred under hypnosis was accompanied by a distinct change in the brain. It is difficult to imagine that such changes in the brain could be due to efforts to enact a role.

Anesthetic Depth

Although daydreaming is quite common, the next example of a change in consciousness is quite uncommon.

Consider the following: Jeanette awoke on the operating table, her hernia surgery still underway. She could hear her surgeon discussing the shape of her breasts and body while experiencing "a blow torch in my stomach…every tissue tearing like a piece of paper" (Willwerth, 1997). Paralyzed from the muscle relaxants administered along with her anesthesia, she was powerless to tell her surgeons that she was suffering the humiliation of their conversation along with the agony of the procedure they were performing. "You're screaming as loud as you can inside your head. It's like being raped and buried alive" (Willwerth, 1997). This story may read like the opening of a horror film but it is a reality for what some estimate to be as many as 40,000 surgery patients a year (Beyea, 2005). Paradoxically called "anesthetic awareness," it reportedly occurs in 1 of every 1,000 surgical patients (Molyneux, 2000). The paradox lies in the very definition of anesthesia. Prys-Roberts (1987) defined anesthesia as a state of drug-induced unconsciousness in which the patient neither perceives nor recalls a noxious stimulus. By definition, therefore, anesthesia assumes a lack of awareness. Indeed, this paradox points to a critical point in understanding anesthesia and consciousness: As we have noted, states of consciousness represent a continuum, not an all-or-nothing state of consciousness or unconsciousness (except perhaps at extreme ends of the continuum).

Early investigations of the value of inhaled anesthetics described anesthesia as a series of stages characterized by levels of analgesia and lack of awareness (Guedel, 1937). As anesthetic awareness indicates, however, determining accurately the level of consciousness during surgery is easier said than done. Historically, physicians have relied on measurement of respiratory rate, blood pressure, and other autonomic responses as measures of anesthetic depth. It may seem like an odd choice when one considers that anesthetics suppress activity of the reticular formation, thalamus, and the cerebral cortex, considered to be the most important area involved in awareness (Steriade, Amzica, & Contreras, 1994). Though researchers consider autonomic responses to be an indirect measure of anesthetic perfusion in the brain, we need a more direct measure allowing for the quantification and prediction of awareness (Daunderer & Schwender, 2001).

To that end, researchers have developed a number of techniques over the past decade to more directly monitor awareness during anesthesia. One of the first, bispectral analysis, emerged as a means of quantifying awareness in patients. Bispectral analysis operates by converting an electroencephalographic signal into a power spectral analysis (March & Muir, 2005). The measure, abbreviated BIS, is derived from three factors measured by the EEG. The first is the extent to which EEG waveforms show biocoherence, which is the level of variability displayed in brain waves. The less variable the brain waves produced, the more biocoherence, and vice versa. Biocoherence is generally seen as a measure of slowed brain activity (Andreassi, 1995). Indeed, the amount of biocoherence increases with increases in anesthetic depth (March & Muir, 2005). The second factor in the BIS calculation is the amount of power in the delta activity versus that in the beta activity. Delta waves generally are characteristic of a lower level of consciousness, whereas beta waves typically reflect increased levels of brain activity, like that which might occur during a problem-solving task (Andreassi, 1995). As anesthetic depth increases, the brain activity begins to be dominated by more and more delta activity at the expense of beta (March & Muir, 2005). The third criterion for BIS calculation is the proportion of isoelectric activity. Isoelectric activity is characteristic of no brain activity. As one would expect, therefore, as anesthetic depth increases, so does the proportion of isoelectric activity (March & Muir, 2005). The BIS algorithm results in a number ranging from 0 (no brain activity) to 100 (fully awake). Glass and Johansen (1998) suggest that to avoid awareness, values between 40 and 60 BIS should be maintained to achieve sufficient anesthetic depth to prevent awareness. These claims were put to the test in an extensive experiment of 2,463 surgical patients. Myles, Leslie, McNeil, Forbes, and Chan (2004) compared traditional anesthetic monitoring to use of the BIS in preventing awareness of surgical patients. They found a reduced risk of awareness of 82 percent when using BIS monitoring. Of the over 1,200 BIS assigned participants, only 2 had an episode of awareness compared to 11 such cases in the non-BIS assigned condition (Myles et al., 2004).

The value of the BIS demonstrates a cardinal point regarding consciousness. Consciousness is not a binary, all or nothing, phenomena. Indeed, it is a continuum, where varying levels of brain activity manifest similarly in degrees of awareness.

The idea of awaking on the surgical table, unable to protest, is frightening; however, the topic of measuring consciousness has captured the attention of scientists, politicians, and religious leaders alike. What level of consciousness is required to represent human life? When does life end? Who has the right to make that determination? Are people in comas still aware, still conscious?

Disorders of Consciousness

The much-publicized and politicized death of Terri Ann Schiavo thrust issues of consciousness after brain injury back into the public spotlight. Despite proclamations by both sides, the issue is not a simple one. Indeed, assessment of consciousness in brain injury patients has been a source of concern and research by neurologists at least since the 1940s (Bernat, 2006). Researchers have described consciousness as having two separate components: wakefulness and awareness (Plum & Posner, 1980). Wakefulness is largely governed by the reticular formation, a series of chemical networks operating to arouse higher brain function via neurotransmitters such as acetylcholine, and its related thalamic regions (Zeman, 1997). Awareness appears to be largely a function of higher brain structure functioning. The thalamus and its white matter connections to the cerebral cortex appear to be the key structures in the awareness portion of consciousness (Bernat, 2006). Differential damage to one or both of these systems results in various disorders of consciousness.

Jennet and Plum (1972) were the first to describe a disorder of consciousness that is produced by damage to the thalamus, cortex, or its white matter connections—the *persistent vegetative state*—as wakefulness without awareness (Bernat, 2006). Though the patients appear to lack the ability to feel or to perceive themselves or their environment, they do demonstrate a number of behaviors that can lead to confusion regarding this point. Those exhibiting a persistent vegetative state demonstrate a normal sleep-wake cycle with associated eye movements and sexual arousal, the primary difference between this state and coma. Furthermore, they typically exhibit reflexes maintained via the cranial nerves such as blinking, roving eye movements, brief visual pursuit, and nystagmus (Bernat, 2006). Sufferers often demonstrate auditory startle, withdrawal from noxious stimuli, and grimace to pain. Due to the wide range of demonstrated behaviors, it is not surprising that physicians and families alike have demanded additional evidence for the loss of awareness necessary for the diagnosis. Specialists use neuroimaging and electrophysiology to provide further confirmation of this devastating diagnosis. CT and MRI of those diagnosed with a persistent vegetative state show "widespread cortical and thalamic atrophy that increases in severity after months to years" (Bernat, 2006). Further, use of PET scans has shown a decrease in metabolic activity of the cerebral cortex by 40 to 50 percent of normal activity, with damage especially focused in the prefrontal and posterior parietal regions that researchers believe are necessary for attention (Bernat, 2006). Electroencephalography of a vegetative state is typically characterized by slowed background activity, with delta activity that does not react to external stimuli (Bernat, 2006). Reemergence of alpha activity—low-voltage, mixed frequency activity—is typically associated with recovery of awareness (Anch, Browman, Mitler, & Walsh, 1988).

Evidence for some recovery of awareness from diffuse neuronal injury does exist. Patients diagnosed with a minimally conscious state show intermittent, though limited, self-awareness (Bernat, 2006). Patients in a minimally conscious state may potentially demonstrate ability to follow simple commands, make verbalizations, sustain visual pursuit, and demonstrate appropriate emotional reactions (Bernat, 2006). Electroencephalography in these cases is characterized with a nonspecific slowing but contains periods of all normally associated brain activity.

As we've seen, whether it is brain injury or anesthesia, awareness is not an all-or-nothing proposition. Levels of consciousness lie on a continuum that is demonstrated not only in the behavioral repertoire available but also the imaging and electroencephalographic record of brain activity. Alteration in the level of consciousness is not a rare event, however. Indeed, each of us spends an average of eight hours a night in an altered state, asleep.

SLEEP

Connie, a reporter for a local paper, has no trouble falling asleep. In fact, she spends a good portion of her day fighting the urge to close her eyes and take a nap. She retires for bed regularly at ten o'clock at night, falls asleep with ease, but struggles to waken eight hours later when her alarm sounds. How is it a woman who sleeps eight hours a night can still be so tired? To answer this or any question regarding an individual's sleeping habits, one needs two things. The first is an understanding that sleep is not a simple thing. Indeed, there are several different types of sleep, each of which has its own pattern of physiological activity. Second, we need to be able to distinguish between these different types of sleep by measuring the activity of the brain and body during sleep. The recording of physiological activity is called *polysomnography* (*poly* means "many," *somno* means "sleep," and *graph* means "write"). The name reflects that polysomnography detects physiological activity during sleep in many different parts of the body and represents that activity graphically. The body operates electrochemically. Interconnected neurons produce tiny electrical signals, called *action potentials*, which propagate and travel. These action potentials are the means by which neurons carry out the communication necessary to keep the organism running. The signals produced are tiny, ranging from one microvolt to one millivolt (Anch et al., 1988). Bioelectric sensors attached to the sleeper amplify the signals and produce a waveform that can be interpreted to determine the type of sleep the individual is undergoing. Electrodes of this type are placed on the scalp to measure brain activity (EEG-electroencephalogram), the orbital muscles to detect eye movements (EOG-electrooculogram), and on the muscles of the chin, leg, or both to detect neuromuscular disorders (EMG-electromyogram). Sleep researchers may also use electrodes or other measuring devices to assess core body temperature, breathing, heart rate, blood pressure, and so on.

Using polysomnography, researchers and physicians can look at real-time changes in a sleeper's physiology as they happen. This procedure allows researchers to track an individual's sleep pattern. Like waves of light or sound, brain waves can be described using two primary characteristics: amplitude (or voltage, typically measured in microvolts, μV) and frequency (measured in cycles per second; Andreassi, 1995). Some patterns of voltage and frequency are so common that they are given names. *Alpha waves* are characterized by a frequency of 8 to 12 Hz with a magnitude of approximately 20 to 60 μV and are primarily seen over the occipital cortex. Alpha waves are very common in a wakeful, relaxed brain. When the brain begins to become more active it typically produces what are called *Beta waves*. Beta waves have a frequency that is nearly double that of alpha waves and a voltage that ranges between 2 and 20 μV. The *Delta wave* is much different; it is a high-amplitude, low-frequency wave. With a frequency ranging between a much slower 1 to 2 Hz, it is the lowest frequency wave produced. The amplitude can be as large as 200 μV. Delta waves generally indicate that a person is in deep sleep; they are also associated with certain sleep disorders such as sleepwalking and night (or sleep) terrors. *Theta waves* are an additional wave used to discriminate the type of sleep being experienced. Theta waves occur at a frequency of approximately 4 to 7 Hz and at an amplitude ranging between 20 and 100 μV (Andreassi, 1995).

By detecting changes in these patterns of brain activity as well as observing additional changes in the behavior of the organism during sleep, researchers have been able to divide sleep into a series of stages.

Stage 1 Sleep

Not surprisingly, the first stage of sleep is a transitional one. In fact, some debate still exists as to whether Stage 1 is actually sleep (Anch et al., 1988). The EEG during Stage 1 sleep is typified by a mixture of low-amplitude waves including a large amount of alpha activity. It is not uncommon for individuals in this stage to experience an altered state of consciousness often referred to as *hypnogogic hallucinations*. Sleep-induced sensory distortions, they are often experienced as floating, falling, or as a presence looming over the sleeper. Some researchers have suggested that hallucinations of this type are to blame for legends such as the incubus, a nighttime demon that was believed to mount the chest of a sleeper.

Stage 2 Sleep

The transition from waking to deep sleep continues in Stage 2 sleep. EEG activity in the first and second stages is similar, save two oddities that occur only in Stage 2 sleep. EEG events known as *sleep spindles* and *K complex* are unique to Stage 2 sleep. Sleep spindles are defined as bursts of activity between 12 to 14 Hz that last at least half of a second. K complex, also half of a second in duration, has a large negative component followed by a positive deflection (Anch et al., 1988).

Slow Wave Sleep

Stages 3 and 4 often are combined under the name *slow-wave sleep*— a sensible name, as both stages are characterized by high-amplitude, low-frequency delta waves. The two stages differ only in the proportion of delta waves that manifest. Stage 3 sleep is defined as being < 50 percent delta, whereas Stage 4 is > 50 percent delta. As noted earlier, delta wave activity indicates a low level of brain activity, and indeed, Stage 4 is the deepest stage of sleep.

REM Sleep

Stage 4 may be the deepest stage of sleep but it isn't the final stage experienced by a sleeper. That distinction goes to REM, or rapid eye movement sleep. The EEG of REM sleep differs greatly from the preceding slow wave sleep and appears more like mixed low-amplitude activity associated with Stage 1 sleep. So how does one distinguish between REM and Stage 1? The EEG provides one clue in the form of a "saw tooth" pattern of activity that exists in REM but not in Stage 1. The second distinction lies in the name of the stage itself, eye movements (Siegel, 2005). Whereas the first four stages of sleep show little eye muscle activity, EOG measurement during REM shows episodes of darting, saccadic eye movements. Measurement of EMG is also helpful in distinguishing REM from the other stages of sleep. During REM sleep, EMG is at much lower levels than during Stages 1, 2, 3, and 4. What is the source for this difference? Muscle immobility. Motor activity is suppressed during REM, leaving us nearly paralyzed (Siegel, 2005). Additionally, REM sleep is the stage of sleep in which dreaming most often takes place. These three significant differences between REM and the previous stages led to describing Stages 1, 2, 3, and 4 as non-REM (NREM) sleep.

Why Sleep?

It seems a sensible question. Why would we spend one-third of our lives in bed? Surely there is a better use of that time? So what function does sleep serve? One way researchers and physicians have sought to answer this question is by studying what happens when a person doesn't sleep. The question is a practical one as well. It was recently reported that 20 percent of Americans sleep fewer than 6.5 hours per night (Dinges, Rogers, & Baynard, 2006). What's more, a telephone survey of 1,506 adults sponsored by the National Sleep Foundation indicates that they sleep less on weekdays than on weekends. Forty percent reported sleeping fewer than seven hours per night on weekdays; 25 percent got fewer than seven hours of sleep on weekends. The difference in weekday and weekend sleep suggests that they are attempting to overcome a "sleep debt" (National Sleep Foundation, 2005). Research suggests that chronic sleep deprivation of this type can produce significant effects on an individual. Noted sleep researcher James Maas (1998) writes, "The third of your life that you should spend sleeping has profound effects on the other two thirds of your life, in terms of alertness, energy, mood, body weight, perception, memory, thinking, reaction time, productivity, performance, communication skills, creativity, safety, and good health" (p. 6). One large-scale experiment investigated the cognitive effects of sleep deprivation (Van Dongen, Maislin, & Mullington, 2003). Truck drivers were assigned to a week of controlled sleep duration (three, five, seven, or nine hours) each night. Participants displayed a decrease in mean response speed and an increase in the number of lapses in the psychomotor vigilance task (PVT) that worsened throughout the sleep restriction schedule (Van Dongen et al., 2003). Participants assigned to nine hours of sleep showed no decrease in reaction time or increase in lapses. Chronic sleep deprivation clearly has significant affects on vigilance and reaction time. This is especially troubling when you consider that these participants were truck drivers who drive tons of machinery for a living in the lane right next to you and your Dodge Neon. A second study bolsters this concern by reporting that one night of sleep restricted to five hours produced a decrease in performance and an increase in accidents on a driving simulator (Dinges et al., 2006). The effects of chronic sleep loss aren't limited to reaction time, with studies suggesting endocrine, immune, cardiovascular, and metabolic disturbances may also emerge as a result of poor sleep (Dinges et al., 2006).

Sleep loss clearly has profound effects on the individual, leading some researchers to propose a model that draws connections between sleep physiology and cognitive/behavioral abilities during waking. Borbely (1982) proposed the two-process model as a means to explain the function of sleep. The model proposes that sleep has two primary components.

The first component is homeostatic. Homeostasis refers to the means and methods by which the human body attempts to maintain a balance. The model suggests that there is a strain on the body during wakefulness that builds as the waking day goes on. This strain is then quickly lowered during sleep, allowing for the process to repeat itself from zero the following day. Indeed, scientists have recognized a pattern of chemical buildup and breakdown tied to the sleep-wake cycle for many years. One chemical often considered as a possible player in this process is adenosine (Thakkar, Winston, & McCarley, 2003). The body uses adenosine triphosphate (ATP) as a source of energy in all cellular metabolic processes. Metabolism of ATP results in the production of energy and of adenosine as a "waste" product. As a result, adenosine builds up to higher and higher levels during wakefulness, when the body is most active. In fact, researchers have reported that adenosine levels are correlates with perceived sleepiness. During sleep, adenosine levels rapidly decrease and by morning have returned to the levels of the previous morning (Thakkar et al., 2003).

The second component of the model is circadian. In the 1970s the suprachiasmatic nucleus (SCN), a small region of the hypothalamus, was discovered to be the center of circadian timing. Circadian, literally meaning "about a day," refers to the endogenous changes of an organism that occur on an approximate 24-hour cycle. These daily rhythms occur in a number of physiological and neurobehavioral variables such as body temperature, hunger, and, most prominently, the sleep-wake cycle (Czeisler, Buxton, & Khalsa, 2006).

What Dreams May Come

Understanding of the underlying physiology of sleep may be a relatively new achievement but interest in dreams is perhaps as old as written human history. Dreams, perhaps more than any other domain, represent how powerful changes in consciousness can be. Though your body is lying paralyzed in your bed, you believe, however temporarily, that you are skating in Central Park, finishing your Christmas shopping, or perhaps taking that final exam that is coming up next week. The Judeo-Christian tradition is ripe with examples of the regard and curiosity in which dreams have been held throughout human history. The Torah and New Testament both contain multiple stories in which dreams are a means by which God communicates with man (Anch et al., 1988). It is perhaps this tradition that eventually led to dream interpretation as a means of discovering a dream's "true meaning."

Freud and Dreams

For Freud, the dream was not a window to God but instead a window to one's self. Freudian psychology was built upon the notion of the power of the unconscious mind in driving behavior. Hidden from our own awareness lies the id, a powerful driving force of personality that seeks instant gratification. The id, subjugated by the ego during wakefulness, is free of restraint during sleep. This freedom allows the id to fulfill its wishes in the form of dreams. Those wishes are not clearly manifest, however, hidden in symbolism and inner meaning. Whereas the dream as we remember it was called the *manifest content*, it was, for Freud, the deeper hidden meaning that acted as a portent of the id's wishes. With great training a therapist could come to decipher this hidden, *latent content* and from it gain greater understanding of the dreamer (Freud, 1900). Domhoff (2004) reviewed dream content in both children and adults to determine if evidence supports the role of wish fulfillment in dreaming. A longitudinal study failed to support the notion of wish fulfillment in children (Foulkes, 1982). Content analysis of all dreams recorded failed to show any evidence of wish fulfillment by any study participants. A further challenge to the wish fulfillment theory comes from the nightmare dreams that often accompany post-traumatic stress disorder (PTSD). Domhoff points out that these nightmarish dreams are very common following traumatic experiences such as rape, assault, natural disasters, and war. These dreams, clearly not related to wish fulfillment, provide strong evidence against Freud's claims.

Despite little empirical evidence to support Freud's view of dreams, it wasn't until the 1970s that a powerful competing theory emerged (Hobson & McCarley, 1977). Titled the *activation synthesis theory*, Hobson's theory was intended to combat the Freudian notions of hidden dream meanings. The theory holds that dreams are a result of random stimulation of brain activity beginning in the pontine region of the brain stem and ending in the forebrain. Hobson and McCarley (1977; Hobson, 2005) maintain that dreams are simply the side effect of the brain mechanisms involved in REM sleep. The narrative quality of dreams results from the forebrain attempting to make sense out of the random brain activity and not a means by which we can analyze the unconscious mind. According to Hobson and McCarley (1977), it is this process that is responsible for the often nonsensical nature of dreams. The activation synthesis hypothesis has garnered great support from the empirical dream research community, perhaps largely due to its being seen as an answer to Freud. However, recent empirical evidence has begun to bring into question some of the basic tenets of the theory. For instance, researchers have reported for a long time that some forms of dreaming take place outside of REM sleep, calling into question the view of dreaming as a side effect of REM (Foulkes, 1982). Further, the "bizarreness" of dreams is more rare than Hobson's theory would suppose. Using multiple measures of "bizarreness," Strauch and Meier (1996) analyzed 117 dreams. Bizarre elements, such as eccentric actions or unusual structure, were reported as not occurring at all in nearly 25 percent of dreams, whereas 39.3 percent of dreams contained only one bizarre element (Strauch & Meier, 1996).

Though no evidence exists to support dreams as wish fulfillment (Fisher & Greenberg, 1996), evidence is accumulating that dreams may be an expression of one's waking life. Continuity theory postulates that the content of dreams is related to activities and events in an individual's life. Snyder, Tharp, and Scott (1968) analyzed dream content in 58 normal participants. Dream content suggested that the vast majority of dreams (90 percent) involved real-life experiences and everyday concerns. Continuity theory suggests that dreams act less as a means to disguise the wishes of the id and more as a means to express the same kinds of activities, worries, and concerns experienced by sleepers in their waking life.

REFERENCES AND FURTHER READINGS

Anch, A. M., Browman, C. P., Mitler, M. M., & Walsh, J. K. (1988). *Sleep: A scientific perspective.* Englewood Cliffs, NJ: Prentice Hall.

Afifi, A. K., & Bergman, R. A. (2005). *Functional neuroanatomy: Text and atlas*. New York: Lange Medical Books/McGraw-Hill.

Andreassi, J. (1995). *Psychophysiology: Human behavior & physiological response.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Atkinson, R. L., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1990). *Introduction to psychology* (10th ed.). Orlando, FL: Harcourt Brace Jovanovich.

Bernat, J. (2006). Chronic disorders of consciousness. *The Lancet, 367*, 1181–1192.

Beyea, S. C. (2005). Addressing the trauma of anesthesia awareness. *Association of Operating Room Nurses Journal, 81*(3), 603–606.

Borbely, A. A. (1982). A two process model of sleep regulation. *Human Neurobiology, 1*, 195–204.

Bower, B. (2002). Spreading consciousness: Awareness goes global in the brain. *Science News, 162*, 251–252.

Bower, B. (2005). Possible worlds: Imagination gets its due as a real-world thinking tool. *Science News, 167*, 200–201.

Cunningham, S., Scerbo, M. W., & Freeman, F. G. (2000). The electrocortical correlates of daydreaming during vigilance tasks. *Journal of Mental Imagery, 24*(1), 61–72.

Czeisler, C., Buxton, O., & Khalsa, S. (2006). The human circadian timing system and sleep-wake regulation. In M. Kryger, T. Roth, & W. Dement (Eds.), *Principles and practice of sleep medicine* (4th ed., pp. 136–154). New York: Elsevier Saunders.

Daunderer, M., & Schwender, D. (2001). Depth of anesthesia, awareness and EEG. *Anaesthesist, 50*, 231–241.

Davis, S. F., & Palladino, J. J. (2007). *Psychology* (5th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Dinges, D., Rogers, N., & Baynard, M. (2005). Chronic sleep deprivation. In M. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (4th ed., pp. 67–76). New York: Elsevier Saunders.

Domhoff, G. W. (2004). Why did empirical dream researchers reject Freud? A critique of historical claims by Mark Solms. *Dreaming, 14*(1), 3–17.

Fisher, S., & Greenberg, R. (1996). *Freud scientifically appraised*. New York: Wiley.

Foulkes, D. (1982). *Children's dreams.* New York: Wiley.

Freud, S. (1900). *The interpretation of dreams* (J. Crick, Trans.). London: Oxford University Press.

Giambra, L. M. (2000). The temporal setting, emotions and imagery of daydreams: Age changes and age differences from late adolescence to the old-old. *Imagination, Cognition, and Personality, 19*, 367–413.

Glass, P., & Johansen, J. (1998). The bispectral index monitor: Monitoring patient consciousness. *Surgical Service Management, 4*, 50–52.

Guedel, A. E. (1937). *Inhalational anesthesia: A fundamental guide.* New York: Macmillan.

Hilgard, E. R. (1991). Suggestibility and suggestions as related to hypnosis. In J. F. Schumaker (Ed.), *Human suggestibility: Advances in theory, research, and application* (pp. 38–58). New York: Routledge.

Hobson, J. A. (2005). *13 dreams Freud never had: The new mind science*. New York: Pi Press.

Hobson, J. A., & McCarley, R. W. (1977). The brain as a dream state generator: An activation synthesis hypothesis of the dream process. *American Journal of Psychiatry, 132*, 1335–1348.

Hockenbury, D. H., & Hockenbury, S. E. (2006). *Psychology* (4th ed.). New York: Worth.

Honma, H., Masako, K., Fukuda, N., Kobayashi, R., Sakakibara, S, & Koyama, T. (2000). Differences in electroencephalogram power densities between genuine narcolepsy and secondary narcolepsy. *Psychiatry and Clinical Neurosciences, 54*(3), 326–327.

Huettel, S. A., Song, A. W., McCarthy, G. (2004). *Magnetic resonance imaging.* Sunderland, MA: Sinauer Associates.

Jennet, B., & Plum, F. (1972). Persistent vegetative state: A syndrome in search of a name. *The Lancet, 1*, 734–737.

Kihlstrom, J. F. (1985). Hypnosis. *Annual Review of Psychology, 36*, 385–418.

Klinger, E. (1990). *Daydreaming: Using waking fantasy and imagery for self-knowledge and creativity.* Los Angeles: Tarcher.

Kosslyn, S. M., Thompson, W. L., Costantini-Ferrando, M. F., Alpert, N. M., & Spiegel, D. (2000). Hypnotic visual illusion alters color processing in the brain. *American Journal of Psychiatry, 157*, 1279–1284.

Leitenberg, H., & Henning, K. (1995). Sexual fantasy. *Psychological Bulletin, 117*, 469–496.

Maas, J. B. (1998). *Power sleep.* New York: Villard.

March, P., & Muir, W. (2005). Bispectral analysis of the electroencephalogram: A review of its development and use in anesthesia. *Veterinary Anaesthesia and Analgesia, 32*, 241–255.

Molyneux, M. E. (2000). Awake during surgery. *The Lancet, 335*, 671–672.

Montgomery, G. H., DuHamel, K. N., & Redd, W. H. (2000). A meta-analysis of hypnotically induced analgesia: How effective is hypnosis? *International Journal of Clinical and Experimental Hypnosis, 48*, 138–153.

Myles, P. S., Leslie, K., McNeil, J., Forbes, A., & Chan, M. T. V. (2004). Bispectral index monitoring to prevent awareness during anaesthesia: The B-aware randomized controlled trial. *The Lancet, 363*, 1757–1763.

Nash, M. R. (2001, July). The truth and the hype of hypnosis. *Scientific American,* 47–55.

National Institute of Neurological Disorders and Stroke. (2006). *NINDS narcolepsy information page.* Retrieved December 23, 2006, from http://www.ninds.nih.gov/disorders/narcolepsy/htm

National Sleep Foundation. (2005). *Sleep library/Can't sleep? Learn about insomnia.* Retrieved July 30, 2005, from http://www.sleepfoundation.org/sleeplibrary.index

Newman, A. W., & Thompson, J. W., Jr. (2001). The rise and fall of hypnosis in criminal investigation. *Journal of the American Academy of Psychiatry and Law, 29*, 75–84.

Patterson, D. R. (2004). Treating pain with hypnosis. *Current Directions in Psychological Science, 13,* 252–255.

Plum, F., & Posner, J. B. (1980). *The diagnosis of stupor and coma* (3rd ed.). Philadelphia: F. A. Davis Company.

Prys-Roberts, C. (1987). Anesthesia: A practical or impossible construct? *British Journal of Anesthesia, 59*, 1341.

Siegel, J. M. (2005). REM sleep. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (4th ed., pp. 120–135). New York: Elsevier Saunders.

Snyder, F., Tharp, I., & Scott, J. (1968). Psychophysiology. *Phenomenology of REM dreaming, 4*, 375.

Steriade, M., Amzica, F., & Contreras, D. (1994). Cortical and thalamic cellular correlates of electroencephalographic burst-suppression. *Electroencephalographic Clinical Neurophysiology, 90*, 1–16.

Strauch, I., & Meier, B. (1996). *In search of dreams: Results of experimental dream research.* Albany: State University of New York Press.

Streeter, M. (2004). *Hypnosis: Secrets of the mind*. Hauppauge, NY: Barron's.

Thakkar, M., Winston, S., & McCarley, R. (2003). A1 receptor and adenosinergic homeostatic regulation of sleep-wakefulness: Effects of antisense to the A1 receptor in the cholinergic basal forebrain. *The Journal of Neuroscience, 23*(10), 4278–4287.

Van Dongen, H. P. A., Maislin, G., & Mullington, J. M. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep, 26*, 117–126.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*, 158–177.

Willwerth, J. (1997). What's up doc? It's every surgery patient's nightmare: You're awake enough to feel the knife but not enough to scream. *Time*, *150*(8), 96.

Yapko, M. D. (1994). *Suggestions of abuse.* New York: Simon & Schuster.

Zeman, A. (1997). Persistent vegetative state. *The Lancet, 250*(9080), 795–799.

22

Taste

Richard L. Doty

University of Pennsylvania School of Medicine
Hospital of the University of Pennsylvania

The evolution of life on this planet required organisms to sense chemicals suspended or dissolved in water for their survival. Some such chemicals were destructive and required avoidance, whereas others provided nourishment and sustenance. Single-celled organisms such as the bacterium *Escherichia coli* that inhabits our own digestive systems, developed multiple chemical receptors that determined whether they should approach or avoid a given situation. In the case of E. coli, for example, the direction of rotation of the flagella—whip-like appendages that propel them through their environment—is influenced by the type of chemical they encounter. Thus, chemicals important for sustenance produce a counterclockwise rotation of the flagella that facilitates a smooth and somewhat linear swimming path, whereas other, seemingly noxious chemicals induce a clockwise flagellar rotation that produces tumbling and turning away from the source of the stimulus.

Like these simple single-celled organisms, the ability to sense chemicals was critical for the evolution of multicelled organisms, and most developed sophisticated chemosensory systems for that purpose. In both invertebrate and vertebrate species, the sense of taste plays a critical role in determining what materials are eaten or rejected. In mammals, the peripheral elements of this system are the taste buds, bulb-like structures located on the tongue and palate, usually on protuberances termed *papillae*. The human taste bud contains multiple types of sensory receptors that mediate the basic sensations of sweet, sour, bitter, salty, metallic, and perhaps fat and umami. Umami represents taste sensations derived from some amino acids, including monosodium glutamate. The basic taste sensations ensure the energy needs of the organism (sweet, fat, umami), proper electrolyte balance (salty), and avoidance of toxic substances (bitter, sour). Combined with elements of texture, temperature, and smell during eating, such sensations produce the “flavor” of foods. It is important to note that many “tastes” such as those recognized as apple, cherry, chocolate, meat sauce, pizza, strawberry, and lemon actually depend upon the olfactory receptors, not the taste buds. During chewing and swallowing, volatiles arising from the ingested food pass from inside the oral cavity to the olfactory receptors via the opening to the rear of the nasal cavity (i.e., the nasopharynx), producing such sensations. This can be readily appreciated when one holds the nose shut during eating, minimizing the ability of air to freely move from oral cavity into the nasal cavity.

This chapter describes the anatomy, development, and function of the human taste system. The ability to taste begins in the uterus and gradually matures or becomes modified by learning as postnatal life progresses. The appreciation of sweet seems to predominate in the taste world, likely being the first of the basic taste qualities to appear and the last to disappear during the life process.

ANATOMY AND PHYSIOLOGY

Humans possess approximately 7,500 taste buds. The sensory receptor cells within each taste bud are responsive to organic molecules, made up of mostly carbon, hydrogen, and oxygen. Such molecules, termed *tastants*, enter the buds through small openings, termed *taste pores*, located at the surface of the epithelium (Figure 22.1). Multiple buds can be present on a single papilla. Although individual papillae can vary considerably in structure and size, they are typically classified into four major types: *filiform*, *fungiform*, *foliate*, and *circumvallate*. The pointed filiform papillae contain no taste buds, serving mainly to move, and perhaps abrade, food particles during the process of eating. The mushroom-shaped fungiform papillae, most but not all of which harbor taste buds, are found largely on the tip and sides of the tongue (Figure 22.2). The foliate papillae appear as folded ridges along the tongue's margin relatively far back on the tongue. The large circumvallate papillae, which range in number from 6 to 10, resemble flattened hills across the "chevron" of the tongue, each being surrounded by a circular trench or valley.

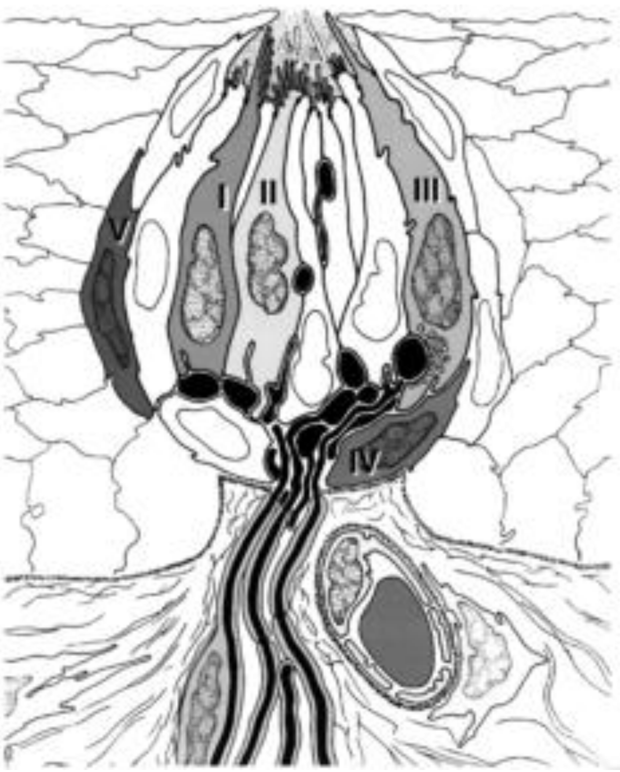

Figure 22.1 Idealized drawing of longitudinal section of mammalian taste bud. Cells of types I, II, and III are elongated and form the sensory epithelium of the bud. These cells have different types of microvillae within the taste pit and may reach the taste pore. Type IV are basal cells, and type V are marginal cells. Synapses are most apparent at the bases of type III cells. The connecting taste nerves have myelin sheaths.

SOURCE: From Witt, Reutter, and Miller, (2005) © 2005 Marcel Dekker, Inc.

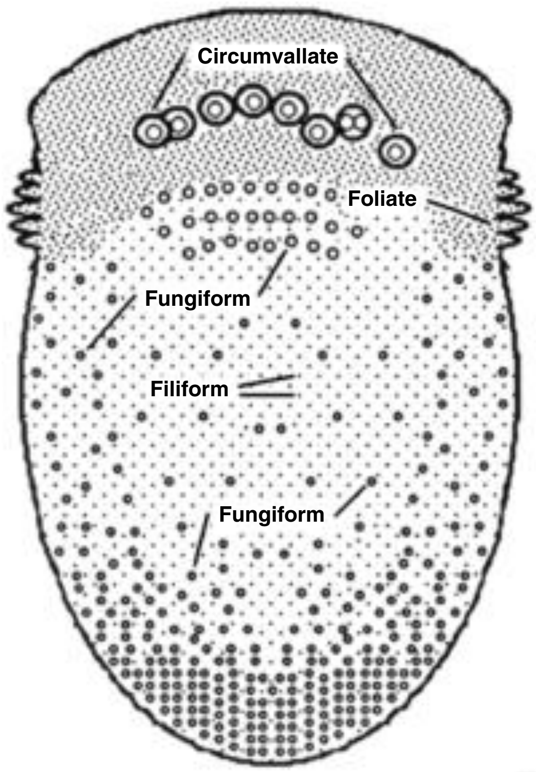

Figure 22.2 Schematic representation of the tongue demonstrating the relative distribution of the four main classes of taste papillae. Note that the fungiform papillae can vary considerably in size, and that they are more dense on the anterior and lateral regions of the tongue.

SOURCE: © 2006 Richard L. Doty.

Like the olfactory receptors, and epithelial cells in general, the neural elements of a taste bud, each of which advances small extensions called *microvilli* toward the taste pore, die and become replaced at various intervals from basal cells. Such receptor cells vary in number from bud to bud, ranging from none at all to well over 100. Unlike the olfactory receptors, the taste receptor cells are not neurons. They form synaptic connections to projection neurons that make up the taste nerves. Depending upon their location, these cells are innervated by one of three sets of paired cranial nerves, namely the facial (CN VII), glossopharyngeal (CN IX), and vagus (CN X) nerves (Figure 22.3).

In humans and most mammals, the chorda tympani division of CN VII innervates the fungiform papillae on the anterior two-thirds of the tongue and the most anterior

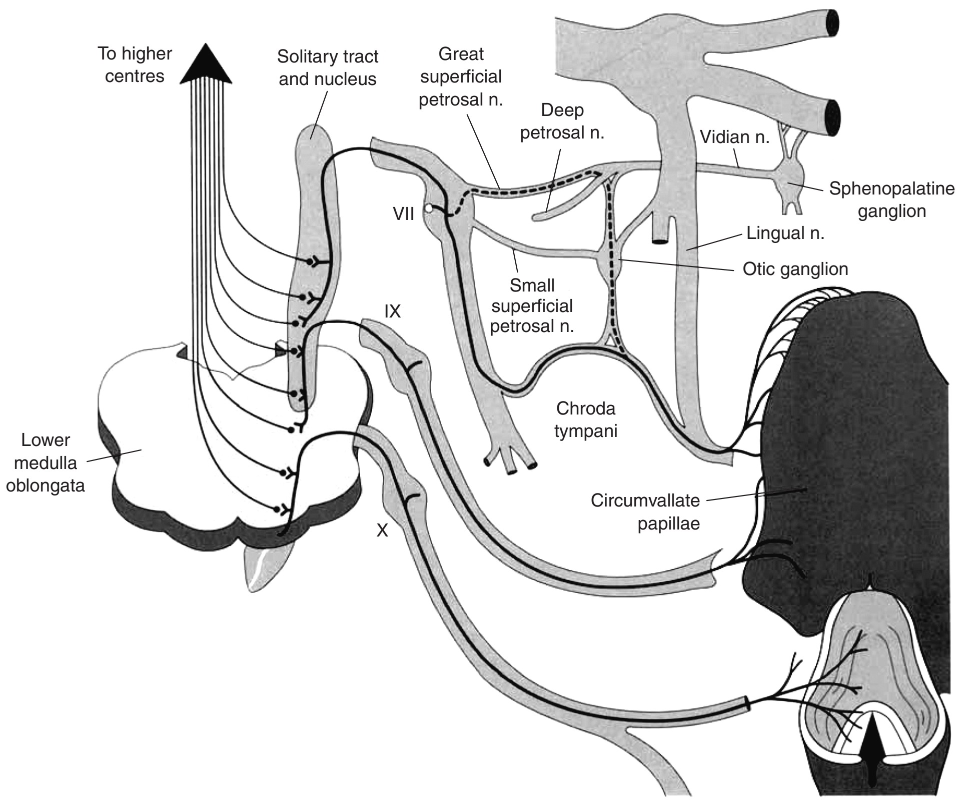

Figure 22.3 Distribution of cranial nerves to gustatory regions. CN VII fibers from the geniculate ganglion innervate taste buds on the anterior portion of the tongue and on the soft palate. CN IX fibers from cell bodies within the petrosal ganglion innervate taste buds on the foliate and circumvallate papillae of the tongue, as well as pharyngeal taste buds. CN X fibers from cell bodies in the nodose ganglion innervate taste buds on the epiglottis, larynx, and esophagus.

SOURCE: Adapted from Netter, 1964. © 2004 Richard L. Doty.

of the laterally located foliate papillae. Because this nerve courses near the tympanic membrane within the middle ear (hence its name), it can be damaged by various pathologies, including ear infections and otosclerosis, as well as by middle ear operations that compromise the eardrum. A common complaint of patients whose chorda tympani has been stretched or otherwise compromised is the presence of chronic "metallic" taste sensations. Damage to this nerve also results in denervation of taste buds and a reduction in the number of taste papillae on the ipsilateral twothirds of the tongue.

The palatine branch of the greater superficial petrosal division of the facial nerve innervates the taste buds within the soft palate (Figure 22.3). The taste buds on the large circumvallate papillae are innervated by CN IX, and those at the base of the tongue, by CN X. Although not involved in taste perception, per se, the trigeminal nerve (CN V) also innervates taste bud regions, as well as most areas of the nasal and oral epithelia, and the teeth. This nerve signals touch, pain, and temperature sensations, and therefore participates in the formation of flavor. Thus, the warmth of coffee and the fizziness of carbonated soft drinks are largely dependent upon the stimulation of fine branches of CN V.

Sensitivity to sweet, sour, bitter, and salty tastants is not uniform across the regions of the tongue. Although diagrams in some children's books and even in some physiology textbooks suggest that different regions of the tongue are responsible for the four basic taste qualities, this is an oversimplification of the facts. In general, the tongue is most sensitive to tastants at its tip, around its margins, and at its base. While it is true that the base of the tongue is more sensitive to bitter tastants, and less sensitive to sweet tastants, than the front of the tongue, both tongue regions can detect bitter and sweet tastants. The relative average sensitivity of tongue regions to the four prototypical taste qualities is shown in Figure 22.4, although significant individual differences are present.

When small tongue regions are compared for their sensitivity between children and adults, children have been found to be more sensitive, presumably reflecting the greater density of receptors within the regions evaluated. However, whole-mouth testing rarely finds children more sensitive than adults. In a recent study, for example, whole-mouth detection thresholds for sucrose, sodium chloride, citric acid, and caffeine were measured for 68 children 6 to 8 years of age and for 61 young adults. Thresholds were not different between girls and either the adult men or the adult women. Boys were less sensitive, on average, than the adult women were to all of these stimuli and less sensitive than the men were to all but caffeine. They were also less sensitive than girls were to sucrose and sodium chloride.

The sensing of chemicals within taste buds is via specialized receptors located within the microvillar membranes of the taste receptor cells. Sweet and umami sensations depend upon a small family of three G-protein-coupled receptors (GPCRs) termed *T1R1, T1R2*, and *T1R3 receptors*. Bitter sensations are mediated by a family of

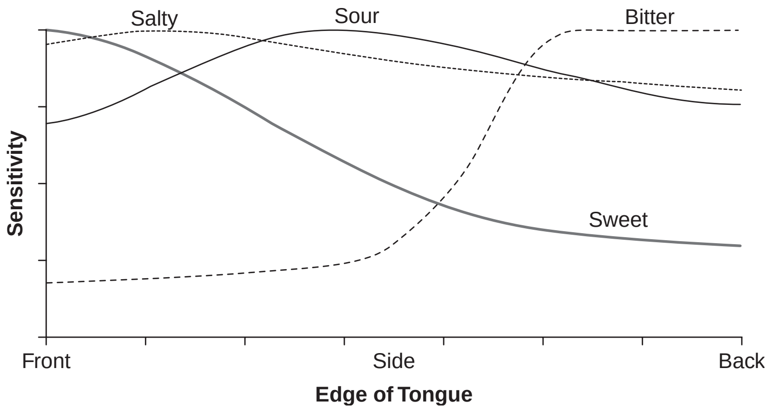

Figure 22.4 Relative sensitivity of the edge of the tongue to the four classic taste qualities. Sensitivity reflects the reciprocal of the threshold value and is plotted as a ratio of maximal sensitivity = 1. Threshold data are from Hänig (1901). Note that all regions of the tongue that were evaluated were responsive to some degree to all stimuli, but that the anterior tongue was most sensitive to sweet, sour, and salty, and least sensitive to bitter. The rear (base) of the tongue was relatively more sensitive to bitter.

SOURCE: Adapted from Boring, 1942.

~30 GPCRs, the T2R receptors, which are expressed on cells different from those that express sweet and umami receptors. T2Rs, which are expressed on the same cells, recognize a wide range of bitter substances, but do not distinguish between them. The salt sensation of sodium chloride is believed to arise from the direct entrance of Na^+ ions into the cells via specialized membrane channels, such as the amiloride-sensitive Na^+ channel. Although a range of receptors has been suggested to mediate sour taste, PKD2L1 has been suggested to be the primary, if not sole, receptor. The deletion of the gene expressing this receptor, a member of the transient receptor potential protein (TRP) family, results in mice unable to detect sour tastants but able to detect sweet, bitter, and salty ones.

The peripheral taste fibers integrate information from multiple taste cells and send this information to the first relay station within the brain, termed the *nucleus of the solitary tract of the brainstem* (Figure 22.3). Although large species differences exist, the nerve fibers that send information from the taste bud to the brainstem can be classified into categories based upon their relative responsiveness to sweet-, sour-, bitter-, and salty-tasting agents. In the hamster, for example, sucrose-best, NaCl-best, and HCl-best fibers have been observed based upon electrophysiological responses. Although fibers are "tuned" for rather specific stimuli, they can nonetheless respond to other stimuli. For example, a few sucrose-best fibers also respond to NaCl and HCl. NaCl-best fibers and HCl-best fibers are less tightly tuned than sucrose-best fibers, with more fibers responding to multiple classes of stimuli.

Taste information is carried from the nucleus of the solitary tract to a taste center within the upper regions of the ventral posterior nuclei of the thalamus via the medial lemniscus, a pathway connecting the brainstem to the thalamus. From here, information is sent to the amygdala and several cortical regions. Two of the cortical regions are the primary somatosensory cortex and the anterior-insular cortex, a region of the frontal cortex located near the anterior temporal cortex. Neurons within these regions respond to taste, touch, and, in some cases, odors.

How does the brain identify the quality of a taste stimulus (i.e., whether it has a sweet, sour, bitter, or salty taste)? As previously noted, receptors are dedicated to chemicals known to elicit these sensations. Although there is clear evidence that at least some taste qualities are mediated by specific receptors, the "taste code" interpreted by the brain is reflected not only in information provided by the specific neurons that are activated (labeled-line theory) but also in the pattern of firing that occurs across taste neurons (across-fiber pattern theory). Intensity or strength of the sensations reflects a combination of the number of taste neurons activated, their identity, and their relative firing rates. Although seemingly simple, the brain must remember what a sweet stimulus, for example, tastes like, and a comparison of information coming from the taste pathways must be made at some point with the remembered sensation to allow for its recognition or identification. Moreover, higher brain centers play a significant role in explaining taste contrasts (e.g., something tasting sweeter after prior experience with a sour stimulus), sensory fatigue, integration of multiple taste sensations, loss of awareness of taste sensations due to inattention, the degree to which a stimulus is like and disliked, and other phenomena similar to those well described for olfaction (see Chapter 24).

TASTE IN THE FETUS

By the 10th week of gestation, the taste bud-containing papillae of the tongue, namely the fungiform, foliate, and circumvallate papillae, have developed. In general, the large circumvallate papillae develop earlier than the fungiform papillae. By 10 to 14 weeks, the taste pores are observed in fetal fungiform papillae, although synaptic connections of the early taste bud primorda to nerve fibers are observed by as early as 8 weeks. Taste buds continue to differentiate and make synaptic connections long after the opening of the taste pores.

The human fetus chews, swallows, and even regurgitates during the second half of gestation. At term, the fetus swallows 70 to 260 ml of fluid per day per kilogram of body weight. Such swallowing is important for fetal amniotic fluid resorption, recirculation of lung and urine fluid volumes, gastrointestinal development, and somatic growth and is likely the sole means by which fetal resorption of the principal amniotic fluid electrolytes is made. Whether the taste system plays a fundamental role in such swallowing is not clear, although the frequency of such swallowing is influenced by tastants introduced into the amniotic fluid of animals.

The nature of taste function in the late-term human fetus can be inferred by testing the ability of premature infants to taste. Taste-induced behavioral responses, including facial expressions seemingly reflecting pleasure or displeasure, are present in premature infants at six to nine months of gestational age. Increases in sucking vigor, reflexive salivation, and, in some cases, retching can be induced in premature babies by placing a single drop of lemon juice into their mouths. Sucking responses are inhibited by low concentrations of quinine, a bitter tastant, whereas sweet-tasting stimuli generally increase sucking frequency. Premature infants, as well as neonates, transiently cease crying when given oral solutions of sucrose or glucose, but not water. Such tastants also promote analgesic-like reactions during heel lance and other invasive procedures. Such responses are taste-bud mediated, since sweet solutions are ineffective in producing analgesia when administered directly into the stomach. The calming effects are induced within seconds of sucrose delivery, well in advance of stomach clearance or absorption.

In general, negative responsiveness to salt ingestion seems to develop between the ages of 4 and 24 months postpartum. Nonetheless, there are reports that variable facial expressions to salt stimuli occur in premature infants. For example, in one study 20 premature infants (1.2–2.9 kg) were orally presented with 0.9 percent NaCl. More than half responded with a rejecting grimace, although 4 readily accepted this solution. In another study, a 1 percent solution of NaCl produced indifference in two-thirds of the premature infants, and rejection in the other third.

In summary, taste buds are functional and capable of conveying at least some types of gustatory information to the central nervous system by the sixth gestational month. Such information is likely available to neural systems that organize salivation, sucking, facial expressions, and other affective behaviors at this early age. In adults, not all regions of the tongue are equally sensitive to tastants, with the rear of the tongue being most sensitive to bitter substances and least sensitive to sweet substances. The perceptual "code" for taste seems to involve a combination of factors, including the specific receptors that are activated and the pattern, type, and frequency of firing of nerve cells that project into the brain stem and, from there, to cortical centers.

Influence of Early Experiences With Tastes on Later Taste Preferences

Exposure of fetuses to electrolyte imbalances and accompanying dehydration appears to alter postnatal taste preferences. In one study, 16-week-old infants of mothers who had experienced morning sickness or had vomited during pregnancy exhibited stronger preferences for 0.1 and 0.2 M NaCl than infants of mothers who had no such experiences. They also ingested larger volumes of the stronger of these solutions. Moreover, the babies of the sick mothers were less likely to express aversive facial reactivity patterns, and more likely to exhibit hedonically positive responses, to the salty solutions. This altered salt preference likely continues into adulthood, since adults whose mothers experienced marked morning sickness exhibited, relative to ones whose mothers did not, (a) greater self-reported use of salt, (b) greater intake of salt in the laboratory, and (c) stronger preferences for salty snack food.

After birth, the taste of the mother's milk, which is usually rich in lactose and other sugars, can be influenced by her diet, thereby influencing suckling responses and, ultimately, the amount and quality of milk that is ingested. For example, some infants breast-feed longer if their mothers are on a flavor-rich diet than if they are on a bland diet. Experience with the ingestant also alters suckling behavior, and in some instances novelty of the taste of the milk will increase the amount ingested. The influences of experience can also play an important role in determining the food preferences of children and teenagers. Children increase their acceptance of a novel food, even food that is initially aversive, after repeated dietary exposure to that food. One study, for example, exposed four- to five-year-olds to either plain tofu or tofu made sweet or salty multiple times over the course of several weeks. When tested months later, the children preferred the tofu with which they had experience. This taste preference did not generalize to other foods of similar color and texture (e.g., ricotta) that were made similarly sweet or salty.

In light of such findings, it is perhaps not surprising that context and culture are important for establishing taste preferences. In a study of first- and second-generation Chinese adolescent immigrant boys to Canada, the second-generation boys and those with more acculturated patterns of language use gave higher hedonic flavor and prestige ratings to dessert, snack, and fast foods, and discriminated better among nutrient-rich and nutrient-poor foods.

HUMAN STUDIES OF BITTER TASTE PERCEPTION

A series of studies in the 1930s highlighted the fact that considerable variation exists in the human population in the ability to detect bitter-tasting agents. Although the whole-mouth sensitivity of humans to most tastants follows a normal distribution in the general population, there are notable exceptions. Among such exceptions are a class of approximately 40 bitter-tasting antithyroid compounds, the most widely studied of which are phenylthiourea (PTC) and 6-n-propylthiouracil (PROP). Nearly all human studies examining the genetics of tasting have focused on these two compounds, whose bitter taste is dependent upon the N – C = S structure within the molecules.

The largely bimodal distribution of sensitivity to PTC/PROP led to the classification of subjects as either “tasters” or “nontasters.” Although the inability to taste these agents was first believed to reflect a simple Mendelian recessive trait, subsequent data contradicted this genetic model, and numerous other models were proposed over the years to explain the inheritance of PTC/PROP taste sensitivity. In 1993 a small region was identified on chromosome 7q that is associated with PTC taste sensitivity. This region contained a single gene that encodes a member of the TAS2R bitter taste receptor family. Five worldwide haplotypes in this gene have been shown to largely explain the bimodal distribution of PTC taste sensitivity, accounting for the inheritance of the classically defined taste insensitivity and for 55 percent to 85 percent of the variance in PTC sensitivity. Importantly, the sensitivity of individuals to PTC/PROP is associated with the number of taste buds, as well as the number of fungiform papillae located on the anterior tongue.

The question arises as to why insensitivity to PTC and related compounds exists in the human population, and why populations in different parts of the world have differing numbers of PTC tasters and nontasters. In one study, for example, only 7 percent of Japanese evaluated were nontasters, compared to 30 percent of Caucasians. Early researchers also found that more PTC nontasters were present in populations with nodular goiters and with cretinism. Derivatives of PTC are known to cause the formation of goiters and to block the synthesis of thyroxine within the thyroid gland. One explanation of why the gene responsible for PTC nontasting is maintained in the gene pool is that those who possess it are protected against malaria. Thus, in a study of Ecuadorian Andean communities in which goiter is endemic, PTC tasters were found not to eat a maize which contains a bitter-tasting goitrogen, and therefore they do not develop goiters. However, these individuals were found more likely to succumb to malaria, unlike their nontasting counterparts.

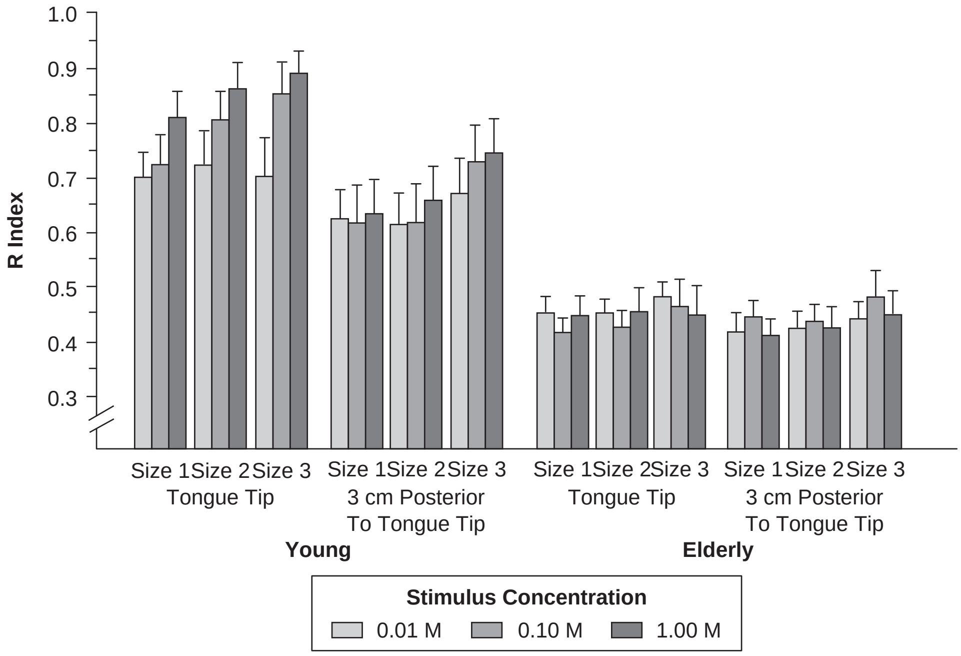

Figure 22.5 Mean (± SEM) sensitivity values obtained from 12 young and 12 elderly subjects for sodium chloride presented to two tongue regions for three stimulation areas (12.5, 25, and 50 mm^2) and three stimulus concentrations. Note that the sensitivity of the older subjects was essentially at chance in all tongue regions and for all stimulus areas assessed. Unlike the young subjects, greater sensitivity was not seen on the tongue tip than on a more posterior tongue site.

SOURCE: Reproduced from Matsuda and Doty, 1995. © 1995 Oxford University Press.

CHANGES IN TASTE WITH ADVANCING AGE

Age-related influences on taste function are well documented. The degree to which age-related effects are observed depends upon the type of testing involved. For example, it is greater when small regions of the tongue are evaluated than when the whole mouth is tested using, for example, “sip and spit” methods. Age-related decreases in taste threshold sensitivity have been reported for such tastants as caffeine, citric acid, hydrochloric acid, magnesium sulfate, propylthiourea, quinine, sodium chloride, sucrose, tartaric acid, and a large number of amino acids. Not all taste qualities exhibit the same degree of age-related loss. Thus, sweet sensations are less ravaged by age than are salty, bitter, and sour sensations. The marked age-related decrement to NaCl seen in small regions of the tongue is shown in Figure 22.5.

As with other sensory systems, a relationship between the number of functioning receptor elements and the system’s sensitivity is present for the sense of taste. Thus, the perceived intensity of tastants presented to localized regions of the anterior tongue is correlated with the number of fungiform papillae and, hence, the number of taste buds within the stimulated regions. Individuals with higher densities of taste buds and fungiform papillae rate the intensity of a number of tastants as more strong than do ones with lower densities of such structures. Although it would seem to follow that age-related declines in the taste function of older persons would relate to decreased numbers of taste buds, there is some controversy as to whether the number of taste buds meaningfully decrease with age. Although age-related losses of taste buds have been reported for human cirmuvallate papillae, most studies of rodent, monkey, and human tongues suggest that taste bud numbers in the anterior and medial lingual regions are little influenced by age. For example, the average percentage of fungiform papillae-containing taste buds in Fischer 344 rats aged 4 to 6 months, 20 to 24 months, and 30 to 37 months was found in one study to be 99.6 percent, 99.3 percent, and 94.7 percent, respectively, differences that were not statistically significant. A human study found no meaningful relationship between age and taste bud densities on either the tip or the mid-region of tongues from young adults (22–36 years, N = 5), middle-aged adults (50–63 years, N = 7), and old adults (70–90 years, N = 6). Given the small sample sizes, however, this research could stand replication with larger numbers of subjects.

Despite the fact that taste bud numbers appear not to be markedly decreased in old rats, electrical responsiveness of the chorda tympani nerve is decreased in such rats to some salts, acids, and sugars. Among the possible explanations of this phenomena are decreased intrinsic reactivity of taste buds to taste solutions, decreased multiple neural innervation of taste buds by some taste fibers, alterations in the general structure of the epithelium (which, for example, might impair the movement of stimulus solution into the taste bud pore), and decreased taste nerve responsiveness, per se. It is also possible that some taste buds, although anatomically present, are not fully functional because of altered turnover time or related metabolic events. Since taste buds function in a complex ionic milieu and are bathed with saliva and other secretory products, changes in such products may also undergo age-related changes. There is suggestion that heightened taste threshold values and flattened suprathreshold psychophysical functions observed in many elderly reflect background tastes noticeable at low, but not at moderate or high, stimulus concentrations. Both neural and oral environment changes (e.g., salivary amount and constituents) could contribute to this noise. Evidence that improved oral hygiene improves taste sensitivity in some elderly persons is in accord with this hypothesis.

Food preferences expressed during childhood, as well as in later life, also reflect genetic determinants that produce considerable variability in acceptance. The best-documented examples of this for children are from studies involving PROP. For example, in one study, five- to seven-year-old children were tested for their sensitivity to PROP. Relative to adults, proportionately fewer "nontasters," defined by threshold sensitivity and suprathreshold ratings of intensity, were found, suggesting that PROP thresholds may rise with age and may partially account for the greater food finickiness observed in many children.

INFLUENCES OF MEDICATIONS ON TASTE PERCEPTION

According to the *Physician's Desk Reference* (PDR), hundreds of medications are associated with taste-related side effects. Terms used to describe such side effects in the PDR include "loss of taste," "altered taste," "ageusia," "taste loss," "dysgeusia," "bad taste," "hypogeusia," "bitter taste," "metallic taste," "unpleasant taste," and "salty taste." Among such medications are widely prescribed ones, including antimicrobials, antifungals, antihypertensives, antihyperlipidemics, and antidepressants. Unfortunately, the true incidence and prevalence of such taste side effects are not known. Most literature reports are anecdotal case reports and only rarely has taste function been measured quantitatively. In some cases, it is unclear if it is the medication or the underlying disorder that is responsible for the taste symptoms, and confounding and interactive influences from other medications may be present, a common situation in the elderly. Since many persons confuse loss of flavor sensations from olfaction with loss of taste, a number of literature reports likely mistake smell losses with taste losses.

That being said, there is convincing evidence that a number of medications alter taste function, most notably ones related to cardiovascular function or its protection. Adverse chemosensory side effects are noted for 70 percent of the antihyperlipidemic drugs listed in the PDR. In a placebo-controlled study of Lipitor (atorvastatin calcium), side effects of altered taste and loss of taste were not uncommon. Similar side effects were found in clinical trials of Baycol, Lescol (fluvastatin), Provachol (pravastatin), Mevacor (lovastatin), and Zocor (simvastatin). Over a third of the antihypertensive drugs listed in the PDR reportedly have adverse taste side effects, including calcium channel blockers, diuretics (e.g., amiloride), and angiotensin-converting enzyme (ACE) inhibitors. ACE inhibitors block the enzyme that converts angiotensin I to angiotensin II, a potent vasoconstrictor that raises blood pressure, and decrease inactivation of bradykinin, a potent vasodilator. Captopril, the first orally active ACE inhibitor, is more frequently associated than any other ACE inhibitor with complaints of ageusia, metallic taste, and taste distortion. This drug can make sweet-tasting foods taste salty, and can produce chronic bitter or salty sensations, presumably by directly altering ion channels. Drug discontinuance usually reverses the taste disturbance within a few months.

Another well-documented drug that severely alters taste function is the antifungal agent, terbinifine (Lamisil). According to PDR, 2.8 percent of 465 individuals who were taking this drug in clinical trials experienced adverse taste effects, as compared with 0.7 percent of a group of 137 persons taking a placebo. Since quantitative assessment of taste function has only rarely been performed,

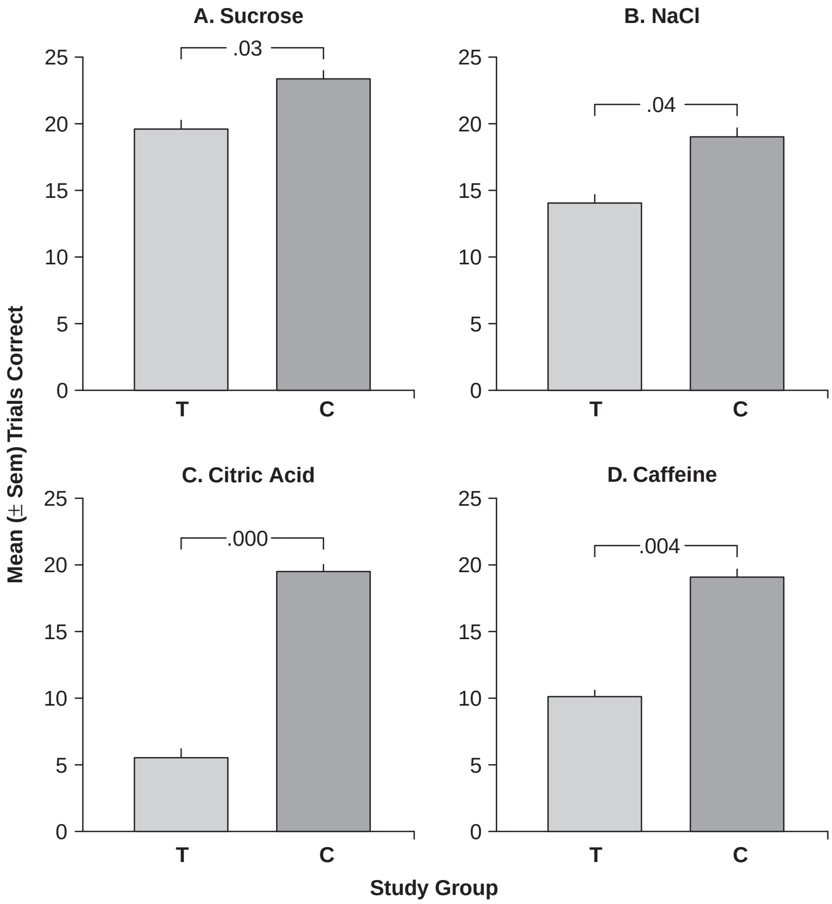

Figure 22.6 Mean (SEM) number of trials correct for the identification of sweet-, sour-, bitter-, and salty-tasting substances in six patients complaining of taste deficits following terbinafine usage (T) and in six matched controls (C). The P values reflect the main effects of group in separate analyses of variance performed for each tastant.

SOURCE: © 2005 Richard L. Doty.

it is likely that a much larger number of persons taking this medication have taste deficits. The subjective taste problems resolve over the course of a few weeks to several months after discontinuation of terbinafine, although long-lasting cases have been reported and no long-term study has been performed employing quantitative taste tests. The reason for the loss of taste is unclear, although inhibition of cytochrome p-450-dependent enzymes at the level of the receptors has been suggested as one mechanism. As shown in Figure 22.6, this medication reduces, in patients presenting to our center with complaints of terbinifine-induced altered taste function, the ability to identify sweet-, sour-, bitter-, and salty-tasting stimuli, with a greater influence on sour- and bitter-tasting ones.

SUMMARY

The sense of taste largely evolved to provide organisms with a means to determine the safety and nutritional value of substances they ingest. In humans, taste preferences, which appear to have innate elements, can be greatly modified by experience, even prior to birth. As a result of such experience and underlying differences in genetic makeup, adults exhibit a wide range of taste preferences and predilections. Human taste buds, the frontline sensors of the taste system, contain multiple types of sensory receptors that mediate basic taste sensations and largely determine, along with the receptors of the olfactory system, the flavor of foods. Unlike the olfactory system, multiple paired cranial nerves subserve taste—namely, the facial (CN VII), glossopharyngeal (CN IX), and vagus (CN X) nerves.

The sensitivity of the tongue to sweet, sour, bitter, salty, and other taste qualities is not uniform across its surface. It is most sensitive to sweet and salty substances in the front, and bitter and sour ones in the back and sides, although all three (front, side, and back) tongue regions can detect, to some degree, all of the major sensory qualities. The middle of the tongue has the least number of taste buds and, therefore, is least sensitive to tastants. Individual taste buds have varying numbers of receptor cells sensitive to the basic taste qualities, with some being sensitive to one, others to only a few, and still others to all basic taste qualities. Sweet and umami tastes are mediated by a small family of G-protein-coupled receptors, whereas bitter sensations are mediated by a family of approximately 30 such receptors. The latter receptors are not found on the same cells as the sweet and umami receptors. Responses to most salty-tasting agents likely reflect the entry of Na^+ ions into the cells via specialized membrane channels. One type of receptor has been recently identified that seems to mediate most, if not all, sour tastes.

After activating taste receptor cells, the taste information is carried from the taste buds to higher brain regions via pathways that are dedicated, to a large but not total degree, to a given class of taste stimuli (e.g., sweet-tasting agents). Some such cells are more tightly tuned to a given class of stimuli than other cells, implying that both “labeled line” and “cross-fiber” processes are involved in the coding of information to be sent to higher brain regions for interpretation. The brain identifies tastes in cortical regions where the incoming sensation can be compared to memory stores, thereby allowing a given taste quality to be recognized.

A number of factors influence the ability to taste, including genetics, age, experience, health, and medications that control microbes, hypertension, depression, and cholesterol metabolism. Sensitivity to a number of bitter-tasting compounds is genetically determined and varies considerably among human populations. Studies of Ecuadorian Andean communities find that persons sensitive to bitter tastants are less likely to ingest toxicants that cause thyroid problems than ones who are insensitive to such tastants. However, such persons are more likely than nontasters to succumb to malaria, likely explaining why this genetic dimorphism is maintained in the gene pool.

REFERENCES AND FURTHER READINGS

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century-Crofts.

Chandrashekar, J., Hoon, M. A., Ryba, N. J. P., & Zuker, C. S. (2006). The receptors and cells for mammalian taste. *Nature, 444*, 288–298.

De Graaf, C., & Zandstra, E. H. (1999). Sweetness, intensity, and pleasantness in children, adolescents, and adults. *Physiology & Behavior, 67,* 513–520.

Doty, R. L., Bagla, R., Morgenson, M., & Mirza, N. (2001). NaCl thresholds: Relationship to anterior tongue locus, area of stimulation, and number of fungiform papillae. *Physiology & Behavior, 72,* 373–378.

Doty, R. L., & Haxel, B. R. (2005). Objective assessment of terbinafine-induced taste loss. *Laryngoscope, 115,* 2035–2037.

Ganchrow, J. R., & Mennella, J. A. (2003). The ontogeny of human flavor perception. In R. L. Doty (Ed.), *Handbook of olfaction and gustation* (2nd ed., pp. 823–846). New York: Marcel Dekker.

Green, L. S. (1974). Physical growth and development, neurological maturation, and behavioral functioning in two Ecuadorian Andean communities in which goiter is endemic. II. PTC taste sensitivity and neurological maturation. *American Journal of Physical Anthropology, 41,* 139–151.

Hänig, D. P. (1901). Zur Psychophysik des Geschmachssinns. *Philosophische Studien, 17,* 576–623.

James, C. E., Laing, D. G., & Oram, N. (1997). A comparison of the ability of 8–9-year-old children and adults to detect taste stimuli. *Physiology & Behavior, 62,* 193–197.

Kim, U.-K., Jorgenson, E., Coon, H., Leppert, M., Risch, N., & Drayna, D. (2003). Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science, 299*, 1221–1225.

Matsuda, T., & Doty, R. L. (1995). Age-related taste sensitivity to NaCl: Relationship to tongue locus and stimulation area. *Chemical Senses, 20,* 283–290.

Mennella, J. A., Pepino, M. Y., & Reed, D. R. (2005). Genetic and environmental determinants of bitter perception and sweet preferences. *Pediatrics, 115,* 216–222.

Mistretta, C. M., & Hill, D. L. (2003). Development of the taste system: Basic neurobiology. In R. L. Doty (Ed.), *Handbook of olfaction and gustation* (2nd ed., pp. 759–782). New York: Marcel Dekker.

Netter, F. H. (1964). *The CIBA collection of medical illustrations, Vol. 1 Nervous system*. New York: Ciba Pharmaceutical Co.

Physicians' desk reference. (2006). Montvale, NJ: Medical Economics Company.

Ross, M. G., & Nijland, M. J. M. (1998). Development of ingestive behavior. *American Journal of Physiology (Regulatory Integrative Comparative Physiology, 43*), *274,* R879–R893.

Temple, E. C., Hutchinson, I., Laing, D. G., & Jinks, A. L. (2002). Taste development: differential growth rates of tongue regions in humans. *Developmental Brain Research, 135,* 65–70.

Witt, M., Reutter, K., & Miller, I. J., Jr. (2005). Morphology of the peripheral taste system. In R. L. Doty (Ed.), *Handbook of olfaction and gustation* (2nd ed., pp. 651–677). New York: Marcel Dekker.

23

VISION

GEORGE MATHER
University of Sussex, United Kingdom

The physical properties of light make it an ideal source of information about the external world. Light is a form of energy emitted by the sun and artificial sources, such as lamps, and travels virtually instantaneously in straight lines, or rays. When light rays encounter an obstacle they may be transmitted, absorbed, or reflected. All three behaviors are important for vision:

1. During transmission through translucent objects, the rays may change direction (refraction), a crucial property of lenses.
2. Absorption by photoreceptors in the eye is necessary for the process of vision to begin.
3. Reflection from an opaque surface provides visual information about its properties—its material substrate, shape, and position.

Paradoxically, light behaves as a particle as well as a ray. Light particles (photons) have a vibration frequency, or wavelength, which is important for color vision.

The eye is an optical instrument similar to a simple camera. A bundle of light rays strikes the cornea (transparent front surface of the eye) and enters the eye through a small aperture (the pupil). The rays pass through a lens and strike the network of cells lining the inner surface of the eyeball (the retina). Due to refraction of the rays by the cornea and the lens, an image of the external world is formed on the retina (see Figure 23.1).

Although the retina of the eye contains an image of the world, the image itself is not intended to be seen. We see the external world before our eyes, not the retinal image inside our eyes. Patterns of neural activity initiated by the retinal image lead to conscious visual experiences that are selective, incomplete, and sometimes inaccurate. An artist paints what he perceives, not his retinal image. Nevertheless, the artist's perception is intimately linked to the information available in the visual world, and to the neural apparatus that processes it.

VISION AS INFORMATION PROCESSING

Luminous intensity in the bundle of light rays passing through the lens varies as a function of spatial position, normally specified in terms of x (horizontal position) and y (vertical position), to build a two-dimensional retinal image. In a dynamic natural environment, image intensity also varies as a function of time (t) and light wavelength (λ). One must also take account of the slight difference in viewing position between the two eyes, which can create small differences in retinal luminance, position, and speed as a function of eye-of-origin (e). All the binocular information available for vision can thus be encapsulated in a function containing five parameters: x, y, t, λ, and e. It is the task of the visual system to derive from this retinal luminance function a meaningful representation of the layout and disposition of the surfaces and objects in view. The task is an immensely complex one,

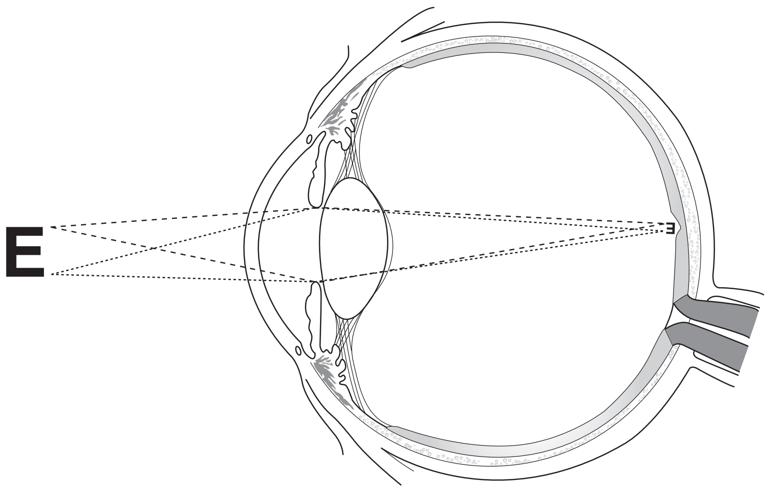

Figure 23.1 Image formation in the human eye.

NOTE: Light rays from an object (dotted lines) enter the eye and are refracted to form an image on the interior surface or retina. The image is inverted and relatively small. A 50mm tall letter 'E,' as in a newspaper headline, casts and image onto the retina that is only 1.5mm tall when viewed at arm's length.

as can be inferred from the limited capabilities of artificially created visual systems in computer vision systems. It is, above all, an information-processing task.

THEORY

Information-Processing Theory: Representation and Computation

Visual stimulation generates electrical activity along the neural pathways from eye to brain. Neuroscientists assume that there is a causal link between this activity and perceptual experience. The neural signals pass through a series of processing stages, embodied in large populations of neurons in the brain, to build internal representations of visible surfaces and objects. Each stage of processing can be conceptualized as an information-processing device. Figure 23.2a shows a simple information-processing device containing three elements: an input, a "black box" processor, and an output. Operations performed by the processor convert the input into the output. The visual system as a whole can be considered to be an information-processing device (Figure 23.2b). The input is the retinal image, and the output is a perceptual event, such as recognition of an object. The black box contains the neural processes that mediate this event. What goes on inside the information-processing black box? Two concepts, representation and computation, are crucial to modern conceptions of visual information processing. The state of one physical system can stand for, or *represent,* the state of another system. For example, the reading on a thermometer may represent ambient temperature, or the display on a vehicle's speedometer may represent its speed. In the case of vision, a specific pattern of neural activity may represent a specific property of the visual world, such as the brightness of a light, the speed of a moving object, or the identity of a face. The information-processing device in Figure 23.2 contains representations created by computations performed on its input. Computation involves the manipulation of quantities and symbols according to a set of rules or algorithms. Any device that takes an input and manipulates it using a set of rules to produce an output can be considered to be a computational device. The visual system is such a device. In order to understand vision, we must understand the representations that are formed in the visual system, and the computations that transform one representation into another.

A large complex problem can be made more tractable by breaking it down into smaller problems. As illustrated in Figure 23.2c, research has shown that the visual system as a whole can be broken down into subsystems. Some subsystems operate in series, so the output of one stage feeds the next stage, and other subsystems operate in parallel. Each subsystem can be considered as a self-contained module of processing devoted to a specific computational task. Signals and computations can be represented in mathematical form as a matrix of numbers, or as a mathematical function. Certain mathematical theories have proved to be especially useful for understanding the computations performed by the visual system.

Linear Systems Theory

One way to characterize the properties of an information-processing device would be to measure its response to all possible inputs. In the case of the visual system, this is an impossible task because the number of possible input images that could impinge on the system is unlimited. If the system can be considered a *linear system,* then it is possible to adopt a simpler approach: measure the responses to a small set of inputs and predict the response to any input. The most commonly used simple visual input is a sinusoidal grating, a regularly repeating pattern of alternating light and dark bars. In order to qualify as a linear system, the visual system's response must obey certain rules. First, if the magnitude of the input changes by a certain factor, then the response must change by the same amount. For example, if the grating doubles in contrast, then the magnitude of response must double. Second, the response to two gratings presented together must be equal to the sum of the responses to the two gratings when presented separately. Third, if the grating shifts in position, the response should not change except for the shift in position. The distance between adjacent bars in the grating (the grating period) should be identical in the input and the output. If all these rules are satisfied, the system qualifies as a shift-invariant linear system.

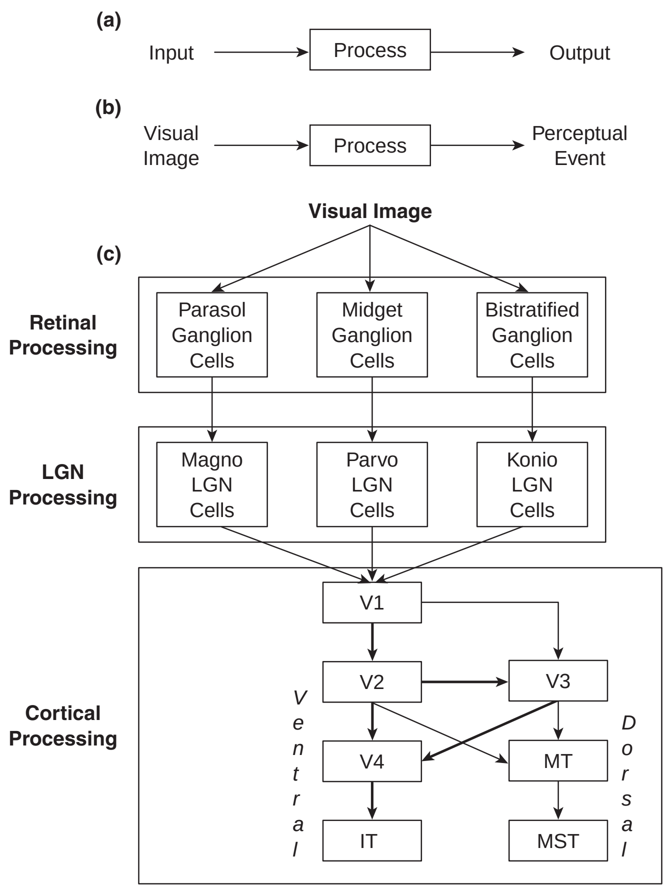

Figure 23.2 Information processing device.

NOTES: (a) A simple information-processing device contains three elements; an input, an output, and a 'black box' processor that converts one into the other. (b) The visual system can be viewed as a complex information-processing device; its input is the retinal image, and its output is a perceptual event. (c) The visual system as a whole can be broken down into component processing modules, each serving a specialized function.

Sinusoidal gratings are special stimuli, according to a mathematical theory called Fourier theory—any complex image can be considered to be a collection of sinusoidal gratings. So once we know how a linear system responds to individual gratings, we can predict its response to any image.

Research has shown that the visual system qualifies as a linear system only in certain restricted conditions. However, the linear systems approach offers a powerful tool for studying the early stages of processing, and for specifying how and when nonlinearities intrude on the system's behavior.

Bayesian Inference

Although linear systems theory is extremely useful for understanding the early stages of visual processing, its utility is more limited when one considers higher levels of processing. Linear systems restrict themselves to the information available in the input, and never extrapolate beyond the available information. But it has become clear that for many visual tasks there is insufficient information available in the input to arrive at a unique solution. For example, the projection of a solid object onto a 2-D image is ambiguous, because many different objects could all potentially project the same image. Does the object in Figure 23.3 have a trapezoidal shape, or is it a rectangle viewed from an oblique angle?

The visual system must go beyond current information and draw inferences based on context, experience, and expectation. Theories of vision have long since acknowledged the need for perceptual inferences, but only relatively recently has another mathematical tool, Bayesian statistics, been brought to bear on the issue. In Bayesian statistics available data is used to infer the probability that a specific hypothesis may be true. A Bayesian inference is called a *posterior probability*—the probability that a particular conclusion is valid (e.g., that the shape is rectangular). It is calculated using two other probability values—a *prior probability* of the likelihood of encountering rectangular shapes in the world, and a *conditional probability* that the image in view could have arisen from a rectangular shape. Bayesian inference encapsulates knowledge of the world in prior probabilities, and sensory data in conditional probabilities. It offers a rigorous mathematical method for evaluating whether the visual system is making optimal use of the information available.

Psychophysical Theory

Visual stimuli are defined in physical terms such as size, intensity, contrast, or duration. They evoke perceptual experiences such as "red," "bright," or "fast." What is the relation between physical stimuli and perceptual experience? Experiments to investigate this relation (known as *psychophysical experiments*) measure the detectability or discriminability of simple visual stimuli as a function of a physical stimulus parameter. Data invariably show a gradual transition between one category of response and another. For example, to measure sensitivity to light the experimenter may manipulate the intensity of a light stimulus, and the task of observers is to report whether they can detect the light or not. A plot of the probability of detection as a function of stimulus intensity (*psychometric function*) should show a gradual shift from no-detection to detection as the stimulus level increases. Classical psychophysical theory (see Chapter 20) explained the smooth function by assuming that the observer's visual system has an internal threshold stimulus value, below which the stimulus is never detected and above which it is always detected. Moment-to-moment fluctuations in this threshold introduce a degree of uncertainty in the observer's responses, creating the smooth curve. Classical theory took no account of other factors that can bear on the observer's decision and add a bias to his or her responses, such as motivation and confidence. Classical theory has therefore been superseded by

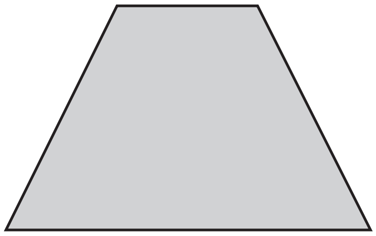

Figure 23.3 What is this shape?

NOTE: Is it a flat trapezoidal panel perpendicular to your line of sight, or a rectangle at an oblique angle to your line of sight? There is insufficient information in the stimulus to make a decision, so the visual system draws inferences based on context, experience, and expectation. Previous experience with obliquely viewed rectangles such as tables and books may lead you to perceive this shape as rectangular.

Signal Detection Theory, according to which participants' decisions are determined jointly by their sensory response and by a bias or tendency to respond in a certain way. Experimental methods used in modern psychophysical research involve techniques that take account of the possibility of bias in observer responses.

METHODS

Psychophysics

Psychophysics is the scientific study of the relation between physical stimuli and perceptual phenomena. Modern empirical research on vision is dominated by psychophysical experiments that measure an observer's ability to detect very weak visual stimuli, such as low-contrast or slowly moving patterns, or to discriminate between two very similar stimuli, such as two patterns differing slightly in contrast or direction. Carefully designed, well-controlled psychophysical experiments can be used to test the predictions of specific computational theories, and can be compared against the results of neurophysiological experiments.

The techniques used in psychophysical experiments were first developed in the late nineteenth century by Gustav Fechner. Psychophysical experiments offer an important perspective on human visual capabilities. However, a more complete understanding of vision also requires that we consider two other perspectives, computational modeling and neurophysiology.

Computational Modeling

As outlined in Figure 23.2, the visual system can be viewed as a hierarchically organized system of information-processing devices. Each device computes an output from a given input, based on a set of transformation rules or algorithms. The precise characteristics of the system can be defined mathematically, and can be implemented on a computer. In principle, then, it should be possible to implement certain aspects of visual function on a computer. Indeed there have been many attempts to build computer vision systems, either to replace a human operator in dangerous environments or as a way of testing theories of human vision. In the latter scenario, the first step in computational modeling is to devise a model that is consistent with available empirical data, and that completes its task successfully. Then the model can be made to generate predictions for the outcome of new experiments, which can be tested psychophysically or neurophysiologically.

Although the theoretical foundations of computational modeling in cognition were laid by the English mathematician Alan Turing in the 1930s, computational modeling become a significant theoretical movement in vision science only during the last twenty years of the 20th century. The change was largely due to the theoretical work of David Marr, a computer scientist who was the first to attempt a comprehensive computational theory of human vision. Though some details of Marr's theories have been rejected on empirical grounds, the general computational strategy is still accepted as a valuable approach to understanding vision.

Neurophysiology

Physiological studies of the functioning of brain pathways mediating vision have predominantly used the single-unit recording technique. A very fine, insulated electrode is inserted into the visual system of an awake, anesthetized animal. Visual stimuli are presented to the animal, and resulting neural activity is picked up by the electrode and recorded on a data logger or computer. The electrode can be inserted at any point in the visual pathway, from retina to cerebral cortex. Recordings reveal the stimulus-dependent response of the cell closest to the electrode.

The single-unit recording technique was developed in the 1950s by Stephen Kuffler and Horace Barlow, who recorded activity in the cat and frog retina, respectively. It was used by David Hubel and Torsten Wiesel in the late 1950s and early 1960s to record activity in cat and monkey cerebral cortex.

During the last 15 years brain imaging has emerged as an important technique for studying the functional organization of the human visual system. The subject is placed bodily in the brain scanner and presented with visual stimuli. In functional MRI scanners, short bursts of radio waves are passed through the brain at many different angles (also see Chapter 16). Neural activity evoked by the stimuli causes tiny changes in the magnetic properties of blood hemoglobin, which are detected and processed to build up a picture of the brain areas activated. The number of functional MRI scanning studies has grown exponentially since the first papers appeared in the early 1990s (exponential growth is typical of scientific research in an expanding field). Brain scans are particularly useful for tracing the brain areas activated by particular kinds of visual stimulus, and can even be used to measure visual adaptation effects.

EMPIRICAL RESULTS AND INFERENCES

Empirical studies of vision fall into camps: neurophysiological experiments and psychophysical experiments. Modern empirical research is tightly coupled to computational theories, so in the following sections empirical results and relevant theory are presented together.

Neurophysiology

Functional Organization

The visual system includes the retinas, the visual pathway connecting the retinas to the brain, and the visual cortex (Figure 23.4). Each retina contains over 100 million photoreceptor cells divided into two classes, rods and cones. Rods are much more numerous than cones but respond only at the low light levels typical of nighttime. Cones respond during daylight, and can be subdivided into three classes according to their response to light wavelengths: short (blue-sensitive), medium (green-sensitive), and long (red-sensitive).

The fibers of approximately one million retinal ganglion cells form the optic nerve, the output from the retina. Optic nerve fibers terminate in the lateral geniculate nuclei (LGN), large bundles of cells lying in the thalamus beneath the cerebral cortex. LGN cells in turn project to the primary receiving area in the visual cortex. The two eyes' fields of view overlap. Half the optic nerve fibers from each eye cross over to the opposite side of the brain (*partial decussation*). So each LGN receives input from both eyes but covers only half of the visual field. This arrangement means that binocular information from the right half of the visual field arrives in the left hemisphere via the left LGN (dotted lines in Figure 23.4), and binocular information from the left half of the visual field arrives in the right hemisphere via the right LGN (solid lines in Figure 23.4). After arrival in primary visual cortex (also known as *striate cortex*), signals are relayed to other cortical areas (known as *extrastriate cortex*).

Viewed as an information-processing system, the visual system can be broken down into a hierarchical series of processing stages. Signals arrive at each stage, are processed by a network of cells, and are then passed on to the next stage. Each cell in any one stage can itself be viewed as a processing device, taking inputs from the previous stage (and from other cells in the same stage) and sending output to the next stage (and to other cells in the same stage).

Figure 23.2c summarizes the main processing stages in the visual system. Each processing stage is represented by a box. The arrows connecting boxes show the main projection paths between stages; the width of each arrow represents the proportion of fibers in each cortical projection pathway. Most ganglion cells can be classified into one of three major types (midget, parasol, and bistratified), which project selectively to cells in the LGN (parvo, magno, and konio, respectively). Primary visual cortex (V1) is the starting point for cortical processing.

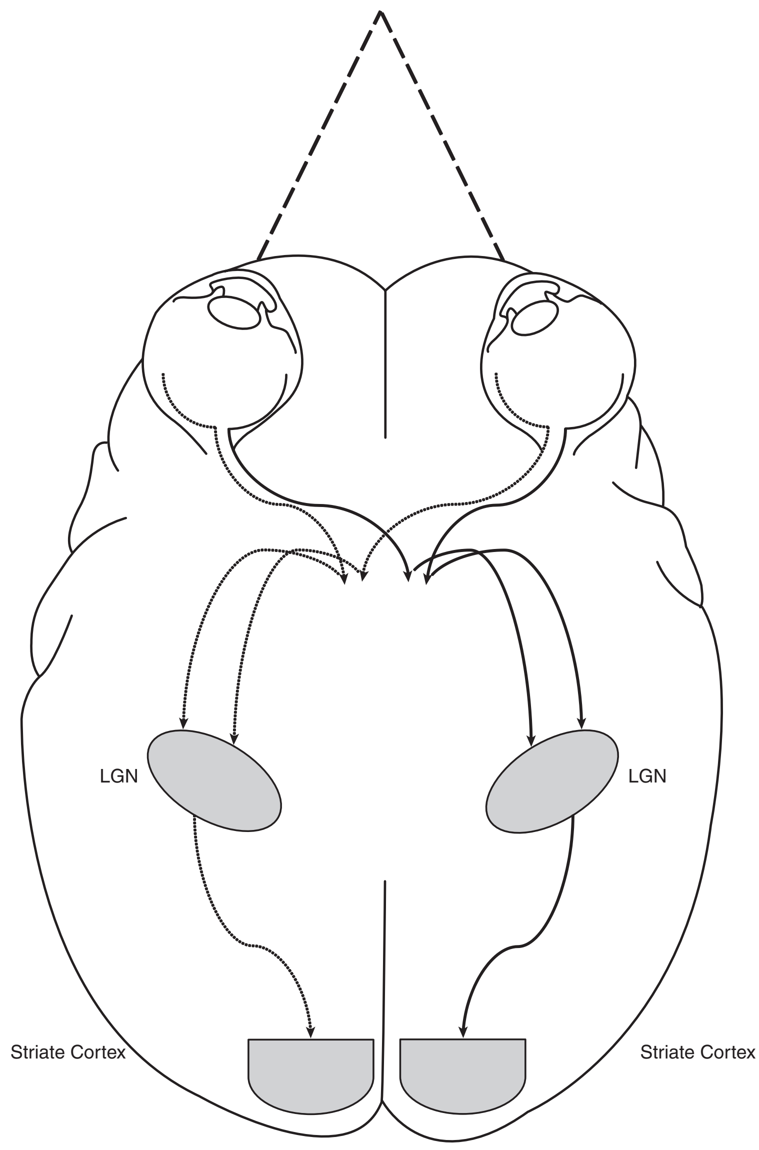

Figure 23.4 Major components of the visual system.

SOURCE: De Valois, Albrecht, and Thorell, 1982.

NOTE: Nerve fibres leave the eye and terminate in two massive bundles of cells in the thalamus called the Lateral Geniculate Nuclei (LGN). Fibres from cells in each LGN then project to the primary receiving area in the cortex, known as striate cortex. Fibres originating in the right half of the visual field converge on the left hemisphere (dotted lines); fibres originating in the left visual field converge on the right hemisphere (solid lines).

Beyond area V2, processing splits into two streams. The ventral stream contains the bulk of cells, and incorporates areas V4 and IT in temporal cortex; the smaller parietal stream includes areas V3, MT, and MST in parietal cortex. Lesion and brain-imaging studies indicate distinct roles for the two streams. Ventral lesions cause impairments in spatial form and color perception. Parietal lesions result in impaired motion direction and speed perception. Brain scans reveal hot spots of activity in ventral cortex, V1, and V2 using color stimuli, and hot spots in parietal cortex while viewing motion stimuli. According to one popular characterization of the different roles of the two streams, the ventral stream specializes in the analysis of color and form, or the "what" properties of the image, and the dorsal stream specializes in the analysis of motion and space, or the "where" properties. According to an alternative characterization, the ventral stream mediates conscious awareness of vision, whereas the dorsal stream mediates unconscious, visually guided action. There is currently no universal agreement on which of these two views offers the best way forward.

Neural Selectivity

Single-unit recordings show that cells in the visual system have a high degree of stimulus specificity. Each cell responds only when a visual stimulus is presented at a specific location on the retina, defining the cell's receptive field. Furthermore, the stimulus must have certain characteristics in order to provoke a significant response from the cell. Stimulus preferences become more elaborate as one advances up the visual system.

Ganglion and LGN cell receptive fields are roughly circular, and most (midget/parvo and parasol/magno) have a center-surround organization; some cells are excited by light falling in the central region and inhibited by light in the surround; other cells have the opposite arrangement. Such spatial opponency means that the cell responds more strongly to patterns of light and dark than to even illumination falling across the entire receptive field. Most ganglion and LGN cells also have some degree of spectral opponency; they are excited by input from one part of the visible spectrum and inhibited by light from another part of the spectrum. Spectrally opponent cells fall into two classes: red versus green (midget/parvo) and blue versus yellow (bistratified/konio).

Receptive fields in visual cortex have more elaborate stimulus preferences. A minority of cortical cell receptive fields (so-called "simple" cells) have elongated, abutting excitatory and inhibitory regions, so the cell responds strongly only when an edge or bar of appropriate size lies at the optimal position and orientation in the receptive field. The majority of cortical cell receptive fields (so-called "complex" cells) do not contain inhibitory and excitatory zones but nevertheless are orientation-selective, so response depends on orientation and size but not position. Many of these cells also respond selectively to the direction of stimulus movement. Some cortical cells respond selectively to wavelength.

Functional Significance of Neural Selectivity

The information present in the visual image varies along five dimensions, as described in the introduction. Each stimulus-selective neuron can be viewed as taking a measurement or sample of the image at a particular location in this five-dimensional space, to provide raw data for later levels of processing. For example, the restricted retinal area of a given cell's receptive field means that the cell measures the image at specific x and y locations, wavelength selectivity samples the wavelength dimension, and direction selectivity provides specific information about image variation over a specific region of space and time. Stimulus-specific neurons thus create a highly efficient representation of the information present in the multidimensional visual image.

Gain Control

Neural activity is metabolically expensive because it consumes energy. Furthermore, natural images have a high degree of statistical redundancy: They are highly predictable in the sense that information at a given location in space and time is likely to be highly correlated with information at a nearby location in space or time. Energy demands and redundancy can be minimized by allowing a given cell's response level to adjust or adapt itself to the prevailing level of stimulation in nearby cells. The process of adjustment or adaptation is known as *gain control*. Neurophysiological studies have shown that the process is implemented by mutual inhibition between nearby cells in the visual cortex. For example, a given motion-selective cell inhibits its neighbors and is in turn inhibited by them. If all cells are highly active, then their responses are redundant. Mutual inhibition serves to reduce their response level or, in other words, to reduce their gain (the rate at which response increases with stimulus level).

Population Coding

Each cell in the visual system responds selectively to a constellation of stimulus attributes—in other words, a specific combination of the five dimensions in the retinal luminance function. But the cell can only offer a one-dimensional output value, its activity level. So it is impossible to deduce the particular mix of values along five dimensions that generated an individual cell's response. For example, it is not possible to infer the orientation of a line in the image from the activity level of an individual orientation selective cell (or a group of cells sharing the same orientation selectivity). The cell's response depends on stimulus size and contrast, and perhaps motion direction, as well as orientation. Stimulus orientation can be inferred by comparing the cell's response level with responses in other cells tuned to different orientations.

Stimulus orientation can be inferred from the preference of the most active cells. This coding principle is known as *population coding*, since it is based on relative activity in a whole population of cells, rather than in single cells.

Psychophysics

Psychophysical research can reveal the functional properties of the visual system, allowing us to enumerate the processing modules involved in visual tasks and the computations performed by the modules. Although links with underlying neurophysiological structures are necessarily tentative, there is often a striking degree of correspondence between psychophysical data and neurophysiological data.

Low-Level Modular Measurement of the Retinal Luminance Function

Sensory adaptation experiments have been especially informative about the nature of low-level processing in human vision. An adaptation experiment involves three phases. In the first phase (preadaptation), the subject's sensitivity is measured to a range of stimuli along some stimulus dimension, such as grating spatial frequency (a measure of bar width) or motion direction. In the second phase (adaptation), the subject is exposed to an intense stimulus at a specific location on the dimension, such as a specific grating frequency or motion direction. Finally, in the test phase (postadaptation), the subject's sensitivity is measured to the same stimuli as used in the preadaptation phase. The difference between pre- and postadaptation sensitivity to each stimulus is taken as a measure of the effect of the adapting stimulus. Results typically show a drop in sensitivity or, in other words, elevated postadaptation thresholds compared to preadaptation thresholds. Modern explanations of the cause of the sensitivity drop link it to gain-control processes identified physiologically and described in the previous section. Crucially, threshold elevation is usually found only for stimuli close to the adaptation stimulus on the relevant stimulus dimension. For example, adaptation is localized to the retinal area exposed to the adapting stimulus and is found only for test stimuli close in spatial frequency, motion direction, and orientation.

Masking has also been used to study low-level visual processing. In this paradigm the subject is required to detect the presence of a target stimulus, such as a grating in the presence of an irrelevant "mask" stimulus. The presence of the mask interferes with performance, but only when the mask is close to the test stimulus along the relevant stimulus dimension. Masking can be explained by the idea that detection is mediated by stimulus selective channels; masks interfere with target detection only when the two stimuli are similar enough to stimulate the same channel.

Psychophysical adaptation and masking studies indicate that early visual processing is served by a bank of filters or channels that each respond to a narrow range of stimuli. These channels send signals to specialized modules that extract information about motion, depth, and color.

Spatiotemporal Channels

Detection of spatial structure such as edges and bars appears to be mediated by a relatively small number of spatial frequency selective channels, perhaps as few as six or eight at any one retinal location. The temporal response of these channels can be examined by measuring observers' sensitivity to flickering lights or gratings using an adaptation or masking paradigm. Results show that flicker detection is mediated by only two or three selective channels; some channels respond over time more quickly than others. Each channel can be described both in terms of its spatial response and its temporal response, namely its *spatiotemporal response*. So at any one retinal location there are perhaps 20 or 30 different spatiotemporal channels providing measurements of the local spatial and temporal structure of the image.

Channels identified psychophysically can be related to the stimulus-specific cells in the visual cortex. For example, Figure 23.5a shows the response selectivity of single cortical cells to grating frequency, obtained from single-cell recordings; Figure 23.5b shows the magnitude of threshold elevation at several different adapting frequencies, obtained psychophysically. The similarity is striking.

The properties of psychophysical channels and of cortical cells can be described quite well using Linear Systems theory. Each channel or cell acts as a linear filter that transmits some stimulus components relatively intact, but removes or filters out other components. Certain aspects of behavior are nonlinear, but these nonlinearities can be identified and accommodated within the Linear Systems framework.

Motion

An important function of spatiotemporal channels is to provide inputs to motion-sensing processes that signal stimulus direction and speed. Motion sensing is one of the most completely understood aspects of low-level visual processing, with sound, direct links between theory, neurophysiology, and psychophysics. Moving objects in the visual field generate patterns of light and dark that drift across the retina, occupying different positions at different times. So in theory, motion can be inferred by correlating two samples of the retinal image that are separated in space and time. A high correlation indicates the presence of motion in the direction of spatial and temporal offset between the sample locations. Well-defined mathematical limits can be placed on the degree to which the two samples can be separated before motion signals become unreliable. Early evidence for neural circuits that perform the necessary spatiotemporal correlation initially came from observations of beetles and flies, in classic experiments performed by Hassenstein and Reichardt in the 1950s. In physiological terms, the correlator proposed by Hassenstein and Reichardt consists of two cells with receptive fields separated by a short distance

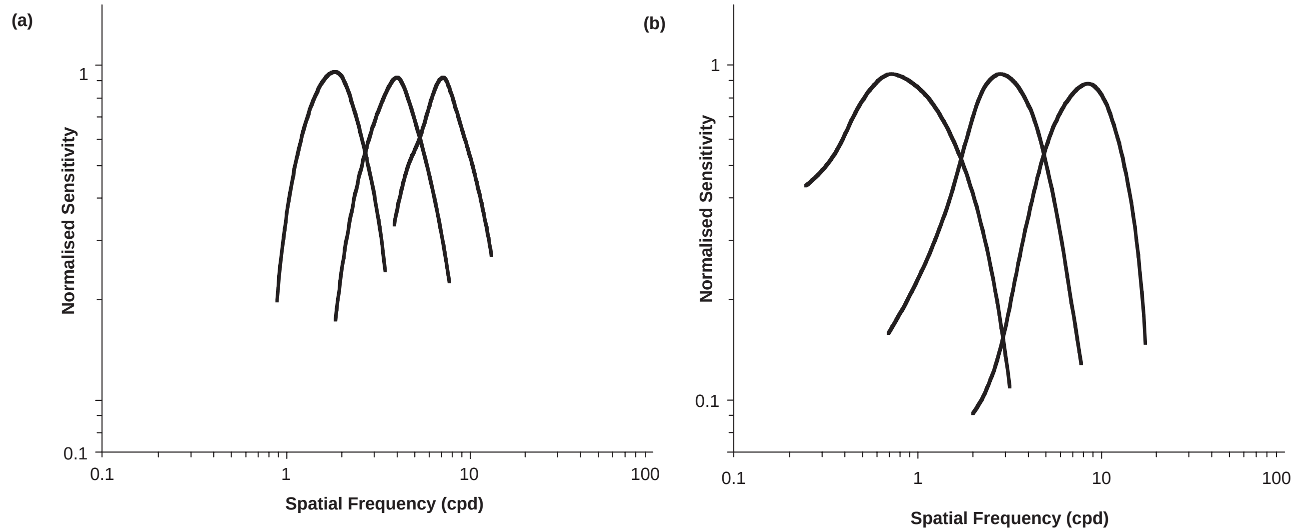

Figure 23.5 Comparison of psychophysical and physiological estimates of spatial frequency selectivity in the visual system.

SOURCE: Wilson, McFarlane, and Phillips, 1983.

NOTES: (a) Relative sensitivity of three individual cortical cells to gratings at different frequency. Each cell responds only to a narrow range of grating frequencies. (b) Relative sensitivity of spatial frequency selective channels in human vision, estimated by adaptation. Three channels are shown.

on the retina, which provide paired inputs to a third cell. The latter cell is the motion sensor. It responds only upon receiving simultaneous signals from the two inputs. One input cell has a faster temporal response than the other. If the difference in response speed matches the time taken for a moving stimulus to traverse the spatial interval between the two input receptive fields, the sensor's response will be direction-selective. It will show a strong response when the stimulus moves in the direction from the slower input receptive field to the faster. Although the faster signal occurs later, by the time it arrives at the sensor it will have caught up with the earlier, slower signal. Movement in the direction from the faster input to the slower produces no response from the sensor because the two signals arrive at different times. In order to detect image motion at many different speeds and in many directions, the visual system must possess sensors with inputs having a range of spatial separations and temporal responses. Physiological studies of many animals, from insects to nonhuman primates, have offered support for the basic correlator theory of motion sensing. In humans, three lines of psychophysical evidence are consistent with the operation of motion-sensing neural circuits.

First, following adaptation to movement in one direction, a subsequently viewed stationary pattern will appear to move in the opposite direction, at least for a short time (the motion aftereffect, or MAE: a demonstration can be viewed at http://www.lifesci.sussex.ac.uk/home/George_Mather/Motion/MAE.HTML). The effect can be explained by population coding of direction in direction-tuned motion sensors: Apparent movement is seen when some sensors respond more than others; adaptation to a given direction lowers the responsiveness of sensors tuned to that direction; and the resulting imbalance in sensor response simulates the pattern of activity that would normally occur in the presence of motion in the opposite direction. So a MAE is experienced.

Second, motion adaptation generates direction-specific threshold elevation. Adaptation to rightward motion, for example, reduces later sensitivity to rightward motion but not to leftward motion. Selective adaptation to direction is a strong indication that direction-tuned channels mediate detection of motion, as explained earlier.

The third line of evidence comes from studies of stroboscopic motion, in which moving patterns are briefly flashed in different spatial positions (as in cine film strips or in stop-frame animation). Stroboscopic motion can mediate realistic impressions of movement, as witnessed by the success of film and television. However, psychophysical studies have found that stroboscopic motion is effective only when the spatial and temporal intervals between animation frames are relatively short. Such short limits are indicative of the involvement of neural correlator circuits. Apparent motion is seen only when the spatial and temporal parameters of the animation match the spatial and temporal intervals built into neural correlators. Psychophysically identified limits on stroboscopic motion perception agree with predictions based on mathematical considerations of sampled motion, and with the properties of cells found in primate cortex.

Although motion sensors play a major role in modern theories of motion perception, they cannot explain all aspects of motion perception. A second, active mechanism is also required, in which motion perception is mediated by spatial shifts in focal attention.

Depth

Most animals have two eyes and receive two images of the world, one in each eye. Human eyes are separated horizontally in the head by an average distance of 6.3 cm. While viewing a natural, three-dimensional world, the small difference in viewpoint causes slight differences between the images received by the eyes. Both eyes face forward and, unless there is a disorder affecting the extraocular muscles, their optical axes converge to allow binocular fixation on a given point of interest in the visual field. The image of a second, nonfixated point in the visual field will be shifted horizontally in one eye's image relative to the other eye's image (horizontal binocular disparity) if that point lies at a difference distance from the viewer than the fixated point. Horizontal binocular disparity offers a precise, quantitative cue to depth. The sign and magnitude of disparity depends on how much nearer or farther away from the observer the nonfixated point lies relative to the fixated point. Disparity increases progressively as the difference between fixated and nonfixated distance increases, though most of the variation occurs within two meters of fixation distance. Random dot stereograms (RDS) were devised by Bela Julesz in the 1960s as a stimulus for studying the use of disparity as a depth cue for human vision. A RDS consists of two fields of random black-white dots, one presented to each eye. The pattern of dots in the two fields is identical, except for a region of dots in one field that is shifted horizontally relative to their position in the other field. The position shift creates horizontal disparity when the RDS is viewed in a stereoscope. Subjects can readily perceive the depth defined by the disparity, proving that disparity alone is a sufficient cue for depth perception. Single-unit recording studies have shown that a class of cells in the primate cortex is disparity selective, and corresponding cells in human cortex almost certainly mediate our ability to perceive depth from disparity. Each disparity selective cell must be binocular; it has a receptive field in each eye. Disparity selectivity can be created simply by shifting the horizontal retinal position of the receptive field in one eye relative to the other eye. Some cells prefer "near" disparity (the nonfixated point is nearer than the fixated point); other cells prefer "far" disparity (the nonfixated point is farther away than the fixated point); a third class of cells prefer zero disparity (the nonfixated point is at the same distance as the fixated point).

A number of other cues also offer information about depth in a viewed scene. A range of monocular visual cues is available (also known as *pictorial cues*), including blur, retinal size, height in the visual field, texture gradient, motion parallax, atmospheric perspective, and interposition:

1. *Blur.* Due to the limited depth-of-focus of the eye, points nearer or farther away than the fixated point may appear blurred in the image. The same effect is seen in photographs.
2. *Retinal size.* The size of an object's image on the retina diminishes as you move farther away from the object.
3. *Height in the visual field.* Objects lying nearby on the ground appear very low down in your field of view; more distant objects occupy higher positions in the image.
4. *Texture gradient.* The width, height, and density of texture elements on a slanted surface all change progressively as your distance away from them increases.
5. *Motion parallax.* The movement in one part of an image relative to another produced by movement through a three-dimensional world such as the relative motion seen from a train window.
6. *Atmospheric perspective*. A reduction in contrast and change in color appearance for distant objects, due to the scattering effect of atmospheric particles such as dust and water.
7. *Interposition*. The partial occlusion of one object by a nearer object.

In addition to a range of visual cues to depth, two nonvisual cues are also available from muscles inside or outside the eye:

1. *Accommodation.* The shape of the eye's lens must change in order to maintain a sharp image of fixated objects at different distances. Muscles inside the eye control lens shape and thus focusing power, so muscle state offers information about viewing distance.
2. *Vergence.* The angle between the optical axes of the eyes must also change with viewing distance in order to maintain binocular fixation. The state of the extraocular muscles controlling vergence angle also provide information about viewing distance.

Most depth cues offer quantitative information about depth that is highly correlated across cues. Although there are many cues available, natural visual tasks such as picking up a teacup, catching a ball, or driving a car require only a single depth estimate for a given object. The object can only be in one place at a time, after all. How are the estimates from different cues combined? It seems that the visual system takes an average of the estimates from different cues, but it weights each cue's contribution to the average. Some cues attract more weight than others. Weights seem to be determined by Bayesian statistics, described earlier. Cues that are more reliable and more likely to be correct are given more weight.

COLOR

As already indicated, human color vision is trichromatic—it is based on responses in three classes of cone photoreceptor distinguished by their different spectral responses: Short wavelength cones are sensitive to light in the blue region of the visible spectrum; medium wavelength cones are sensitive to light in the green region; and long wavelength cones are sensitive to light in the red region. Neurons in the retina, LGN, and cortex code the relative response of the different cone classes by pitting their responses against one another. A given cell may be excited by one cone class, or a combination of classes, and inhibited by a different class or combination of classes. Psychophysically, the crucial aspect of cone response for color appearance is cone excitation ratio—the relative activity across the three cone classes. If two stimuli evoke the same pattern of activity across the three classes, then they will appear identical in color even if the physical wavelengths they emit are different. Such stimuli are called *metamers*. For example, a surface emitting light at a single wavelength of 500 nm may appear green. A second surface emitting light at three wavelengths, 410, 600, and 640 nm, will appear identical in color because it will excite the three cone classes in the same ratio as the single 500 nm light.

Colors are strongly influenced by spatial and temporal context. A green patch surrounded by a red background looks more vivid and saturated than an identical patch surrounded by a neutral colour (*color contrast*); a neutral patch may appear greenish after the observer has adapted to a red patch for a short time beforehand (*color adaptation*). Context effects demonstrate that the color appearance of a surface is governed by a comparison of the cone excitation ratio it generates with the cone excitation ratio generated by its spatial or temporal context. The green patch on a red background looks vivid because its cone excitation ratio shows a greater preponderance of medium-wavelength activity than the excitation ratio generated by the background. A neutral patch appears green after adaptation to red because the reduction in long-wavelength responses due to adaptation biases the cone excitation ratio in favor of the complementary color, green.

Context effects play a vital role in color vision. Color is an unchanging, intrinsic property of an object and potentially useful for discrimination and identification. The distinction between medium- and long-wavelength cones is thought to have evolved to allow our ancestors to distinguish ripeness in fruit and subtle differences in edible leaves. However, the wavelengths given off a surface do not remain constant because they depend partially on the wavelength properties of the illuminating light. Light at midday, for instance, has a high proportion of blue wavelengths, whereas light near dusk contains more long wavelengths. Illumination affects the whole scene by a constant amount. Color contrast and adaptation processes remove constant illumination effects, helping to achieve a high degree of constancy in color appearance.

HIGH-LEVEL VISUAL INFERENCES AND ACTIVE VISION

Low-level modules that encode spatiotemporal frequency, motion, depth, and color provide a piecemeal, localized representation of selected image attributes. Linear Systems Theory has been particularly useful in describing the operation of these modules. Yet surfaces and objects cover much larger areas of the image. The visual system must use grouping operations to build a representation of large-scale structures in the image. High-level grouping processes cannot operate on the basis of linear addition of all local representations. Instead, they require nonlinear interactions such as mutual facilitation and suppression, driven by assumptions about the properties of natural images. These interactions produce a selective representation of the shape and textural properties of surfaces and objects.

GESTALT GROUPING

The earliest demonstrations of such interactions came from Koffka, Kohler, and Wertheimer, the Gestalt psychologists working in Germany in the early twentieth century. They argued that visual system has a propensity to impose certain perceptual groupings on spatial patterns, driven by a set of organizing principles. They saw the most basic principle as the minimum or simplicity principle; pattern elements will be grouped to create the simplest perceptual organization. Their demonstrations show that elements that are visually similar in shape, size, color, or motion and that are close together will be grouped into a perceptual unit. Later work in computational vision by David Marr made use of Gestalt grouping principles to partition the image into regions and shapes that signify the presence of surface and objects. The grouping principles essentially embody assumptions about the visual properties of surfaces and objects. Most objects are made of cohesive, opaque material. So image elements that are close together, or visually similar, or move cohesively, are likely to have originated from the same object. Perceptual grouping processes exploit this assumption.

At a physiological level, perceptual grouping can be implemented using long-range interactions between stimulus-selective neurons, or by reciprocal connections from higher-level processes down to lower-level processes, or both. Cells tuned to the same orientation, for example, whose receptive fields are collinear, may engage in mutual facilitation to reinforce their responses. Cells tuned to dissimilar orientations may mutually suppress each other.

FLEXIBLE PROCESSING

Traditional information-processing systems such as computer programs operate passively. When an input is applied, it triggers a cascade of processing operations that culminates in the production of an output. The computations are predefined to work autonomously and automatically, always producing the same output for a given input. Early computational theories of vision operated in the same way, but recent advances acknowledge that this approach is too simple for human vision. There is considerable evidence that human visual processing is flexible and selective. One major source of flexibility in processing is attention. In the absence of attention, even large-scale changes in the visual world are difficult to detect. Attention is known to modulate the speed of detection and sensitivity to visual stimuli, as well as the depth of adaptation. Neurophysiological studies have found that neuronal responses to attended locations or stimuli are enhanced, while responses to unattended locations or stimuli are suppressed.

Some modern computational approaches favor a hybrid system containing both bottom-up processes, which are applied to the input passively and automatically, and a collection of active, selective processes called *visual routines*. These routines are triggered as and when required for specific tasks, and may include such operations as searching for or counting objects of a certain kind present in the scene.

RECOGNITION

Object representations in the brain are essential for a multitude of everyday visual tasks. The daily routine of eating and drinking uses object representations in a variety of ways:

- For identifying an object as belonging to a particular class of objects, such as a glass
- For discriminating between objects within a class such as your own personal mug
- For interacting with objects such as filling a mug, or picking it up

All of these tasks require that a link be made between a particular visual image and a stored object representation. Objects vary in their intrinsic properties such as shape, size, and surface characteristics. Images of the same object can also vary due to extrinsic factors such as observer viewpoint, lighting direction, and environmental clutter. Theories of object recognition in the human visual system differ in terms of how they deal with extrinsic factors. View-independent theories, such as Marr and Nishihara's theory and Biederman's recognition-by-components theory, attempt to remove extrinsic factors, constructing representations that contain only an object's intrinsic properties. Such theories create a structural description of each object, which is essentially a list of the object's parts and how they link together. View-dependent theories, such as Ullman's alignment theory and Edelman's feature-space theory, incorporate viewpoint effects in the representation. In these theories the representation reflects the characteristics of the object as seen from a specific viewpoint, or small collection of viewpoints. Computational analyses underpin the main proposals of each type of theory, and there is some psychophysical support for each from object recognition experiments. However, neither class of theory can claim unequivocal support. Physiological studies are not definitive on this issue. Some cells in inferotemporal cortex show extreme selectivity for complex patterns such as faces. Their response is not affected by changes in size or location. Other cells respond to specific views of three-dimensional objects. This result does not necessarily favor view-dependent theories because some degree of view dependence would be predicted by both view-dependent and view-independent object recognition theories. According to view-independent theories, structural descriptions are much more readily extracted from some views than from other views.

It may be the case that the two classes of theory are not mutually exclusive. Both view-dependent and view-independent representations are constructed, and have different functions.

SUMMARY

An image of the external world is formed in the eye. Patterns of neural activity triggered by the retinal image travel up the visual pathway from eye to brain, and lead to conscious visual experiences. The visual system and its subsystems can be viewed as information-processing devices; each takes an input, processes it, and produces an output. The system's task is to extract meaningful information from the multidimensional retinal image about the shape, identity, and disposition of the objects in the visual scene. Our understanding of how the visual system achieves this goal has benefited from mathematical and computational analyses of the information available in visual images. Physiological studies show that during early processing stimulus-selective neurons create a highly efficient representation of the spatial and temporal structure in the visual image. Higher-order, more specialized populations of cells extract information about spatial detail, motion, depth, and color. Psychophysical data are consistent with a decomposition of the visual system into specialized modules dedicated to the analysis of specific image attributes. High-level processing achieves order and recognition by applying selective, flexible operations to low-level data that take advantage of assumptions about the characteristics of natural images.

REFERENCES AND FURTHER READINGS

De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22,* 545–559.

Hubel, D. H. (1988). *Eye, brain, and vision*. New York: Scientific American Library.

Landy, M. S., & Movshon, A. J. (Eds.). (1991). *Computational models of visual processing*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Mather, G. (2006). *Foundations of perception*. New York: Psychology Press.

Pasternak, T., Bisley, J. W., & Calkins, D. (2003). Visual information processing in the primate brain. In M. Gallagher & R. J. Nelson (Eds.), *Handbook of psychology* (Vol. 3, pp. 139–185). New York: Wiley.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences, 24,* 295–300.

Wandell, B. A. (1999). Computational neuroimaging of human visual cortex. *Annual Review of Neuroscience, 22*, 145–173.

Wilson, H. R., McFarlane, D., & Phillips, G. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research, 23,* 873–882.

Yantis, S. (Ed.). (2001). *Visual perception: Essential readings.* Philadelphia: Psychology Press.

24

Olfaction

Richard Stevenson
Macquarie University, Sydney, Australia

When you sniff an odor you probably believe that your experience is a direct result of what you are smelling—that is, the *stimulus* dictates what you smell. Many olfactory scientists held this commonsense view until fairly recently, but it is looking increasingly wrong. Rather, what one smells appears to be defined by what the stimulus *reminds* your olfactory system of. If you smell caramel, for example, it smells sweet, yet sweetness is a quality associated with a completely different sensory system—taste. Therefore this experience of a "tasty smell" must be a consequence of memory. This point is made most strikingly when olfactory memory itself is damaged, leaving patients able to smell but unable to experience an odor's unique quality. That is, roses and gasoline, for example, or indeed any odor, all smell alike—a sort of olfactory "grey." In this chapter we examine how the olfactory system works and the crucial role that memory plays in this process, as well as the methods special to olfactory research and the applications that arise from the study of olfaction.

Olfactory capabilities differ. Some animals excel at smelling. Women nearly always outperform men on olfactory tasks. Humans may also possess, along with other animals, two olfactory systems (main and accessory), each with differing capabilities. There are also striking differences in how the main olfactory system processes information at a neural level and in its cognitive processes, when contrasted with the visual and auditory systems. The future direction of olfactory research is exciting; it should help our understanding of the more complex visual and auditory systems that may have evolved from it. It will make a unique contribution to the study of human consciousness. Finally, it has growing practical relevance to the study of human health and disease.

Olfaction is not what you think. Although most of us know that we need eyes to see and ears to hear, most of us do not know that we need a nose to "taste," yet olfaction is central to our enjoyment of eating and drinking. Indeed, most of us probably regard olfaction as nothing more than an occasional sniff of a flower or the heady whiff of garbage or somebody else's unwashed body. But olfaction pervades our life in many ways, betraying whether we are likely to develop Alzheimer's disease or schizophrenia, and indeed, if some are to be believed, influencing whether others find us attractive, as well as allowing us to recover forgotten memories from childhood. However, as interesting as these things are, they are not the reasons that most psychologists choose to study smell. Rather, it is the beguiling mystery of how it works, a conundrum that is still not resolved and at the moment forms the central focus for many psychologists working in this field. This chapter starts by directly addressing this issue and, as you will see, the picture that emerges is indeed bizarre—what you smell is a memory.

THEORY

Function

In both humans and animals, the basic function of the olfactory system is to detect and recognize biologically relevant chemicals (i.e., food, mates, kin, predators). This is no easy matter. The environment is full of chemicals that can be detected by the olfactory system. In addition, nearly all odors—which we perceive as being discrete entities—are in fact combinations of 10s or 100s of different volatile (i.e., smellable) chemicals. For example, chemical analysis of coffee reveals that it has 655 different detectable volatiles; bread, 296; grapefruit, 206; carrot, 95; and even rice has over 100. Yet, in each of these cases, and more generally with most if not all odors, what we *perceive* is a discrete entity or odor object—the smell of coffee, baking bread, or whatever. Theories of olfaction have to account for how we might accomplish this. Before we turn to examine the theories that have been proposed, it is important to consider what we know about the anatomy and sensory physiology of the olfactory system. This physiology constrains any psychological theory, offers its own insights into how the system might work, and presents us with some further problems that any psychological theory must account for. We will concentrate on humans for the moment and defer discussion of the similarities and differences between human and animal olfaction until later (see Comparisons section).

Anatomy and Sensory Physiology

Two major advances have shaped contemporary thinking about olfaction. The first is genetics, derived in no small part from seminal work conducted by Linda Buck and Richard Axel, who won the Nobel Prize for physiology for their study of olfactory receptors. When we detect an odor we do so in two discrete ways. The first is by sniffing, termed *orthonasal olfaction*. The act of sniffing is important, as it directs attention to the stimulus and engages brain areas involved in smell, even if there is no odor present. Damage to brain areas involved in coordinating sniffing can impair an otherwise functional sense of smell, and this is believed to underpin the olfactory deficits observed in patients with Parkinson's disease. The second way we detect odors is not generally associated with smelling and is termed *retronasal olfaction*. When we eat and drink, volatile chemicals are released, and these diffuse up the back of the throat (via the nasopharynx) to the same set of olfactory receptors that are stimulated when we actively sniff something. Most people are not aware of the major role of olfaction in eating and drinking. For example, people who have damaged their olfactory system often report that they "cannot taste"—yet testing reveals an intact sense of taste (sweet, sour, salty, bitter via receptors on the tongue) but a damaged sense of smell. You can readily appreciate this distinction between taste and smell by pinching your nose while eating—all one is left with, in the main, is taste.

The olfactory receptors are located on a sheet of tissue called the *olfactory epithelium*. There is one sheet for the left nostril and one for the right, located just below eye level within the upper part of the nasal cavity. Volatile chemicals diffuse into a thin layer of mucus that covers the olfactory epithelium. This mucus layer is important; if it is abnormally thick (as when you have a head cold), you cannot smell properly. When it is functioning correctly, it probably assists the binding of chemicals to the olfactory receptors. Most important, it allows the body to remove the stimulus (i.e., chemicals) once they have been detected—a problem not faced by the visual or auditory systems. The mucus layer does this by containing proteins that break down or bind to the chemicals and thus lock them into the mucus, which then gradually flows over the olfactory epithelium, down the throat, and into the stomach.

The key player is the olfactory receptor neuron. On its receptor end, which extends into the mucus layer, it expresses G-protein receptors, and it is these that bind with chemicals and lead to depolarization and an action potential. Buck and Axel's work revealed that in mice, there are about 1,000 different sorts of G-protein receptor (just contrast this with the visual system for a moment, with its 4 types of receptor—3 cones and 1 rod). Humans have somewhere between 300 and 500 different sorts of G-protein receptors. These can be arranged into families and subfamilies, based upon their similarity. Similar types of G-protein, not surprisingly, bind to similar types of chemical, but the specific feature of the chemical that they detect (i.e., chemical bond, particular chemical type [aldehyde, alcohol, etc.], chemical shape, etc.) is not well understood. Importantly, each type of G-protein has quite a broad spectrum of sensitivity—it will bind with many types of chemical. This makes it very unlikely that there are specific relations between one receptor and one percept (i.e., there are no "coffee-detecting" receptors).

Each olfactory receptor neuron only expresses one type of G-protein receptor. The arrangement of G-protein receptor types across the olfactory epithelium is not random—certain families of G-protein receptors tend to be located on certain areas of the epithelium—but the functional significance of this is not well understood. Information from *each* type of G-protein receptor (recall that in humans this is somewhere in the range of 300 to 500 types) converges onto clusters of neurons called *glomeruli*, which are in the olfactory bulb. The bulb itself is located underneath the brain and sits a centimeter or so above the olfactory epithelium—one for the left side and one for the right. There are approximately the same numbers of glomeruli as there are G-protein receptor types, and each type of G-protein receptor tends to converge onto just one or two glomeruli. This is significant, as it is highly likely that the information that is passed to the brain (as we will see in a moment) is in the form of a pattern of activity across all of the glomeruli. This pattern has both a spatial (some glomeruli are active, by degree, and some are not) *and* a temporal form (pattern of activity changing over time). This spatial and temporal form is probably unique for every different combination of chemicals that the olfactory system encounters (i.e., coffee will have its own unique spatial and temporal pattern of glomerular activity, as will carrots, white wine, or urine).

The glomerular output then passes largely intact down the lateral olfactory tract into the primary olfactory cortex. This is not true cortex, but paleocortex, with fewer cell layers than found in the neocortex. Nonetheless, it is here that many neuroscientists suspect that three key processes occur. First, if the glomerular output pattern is new, it is learned. Second, where the glomerular output pattern is old, it is recognized via a pattern-matching system. This may then give rise to our conscious experience of smell. Third, the pattern-matching process allows the brain to recognize a combination of chemicals, when other chemicals are simultaneously present in the environment, and to recognize combinations of chemicals when some of their components are missing (a *degraded stimulus*). This pattern-matching process is further assisted by another feature of the primary olfactory cortex, its rapid ability to adapt to chemicals that are present in the environment, thus allowing anything new to be quickly detected. The major downside of the pattern-matching process is that it appears to prevent analysis of the component parts of an odor, making smell a very synthetic, rather than an analytic, sense.

At this point, what we know becomes far sketchier, but a second major advance, namely neuroimaging, has allowed us to identify the areas in the brain that are also involved in odor processing. Whereas our discussion so far has focused on recognizing particular odor objects, there is far more to it. Most odors we smell are perceived in conjunction with other sensory systems (i.e., we *see* the potato chips we are about to eat, we *hear* the crunch in our mouth and *taste* their saltiness). We also respond emotionally to many odors, sometimes very strongly so, especially if they are fecal, rotting, or body odors (the emotion of disgust). Finally, we have some capacity to imagine smells, to name them, and to expect them to have certain properties (e.g., people perceive white wine that has been colored red to have a smell *like* red wine!). We will look now at the areas of the brain that neuroimaging has revealed and the functions they might have.

The primary olfactory cortex has connections to the amygdala. Neuroimaging suggests that the amygdala is involved in our experience of a strong emotional response to certain odors. The primary olfactory cortex also projects to the hypothalamus, and this almost certainly relates to the role of olfaction in hunger and satiety. Most important, the primary olfactory cortex projects to the orbitofrontal cortex. It does so in two ways—first, directly, and second, indirectly, via the thalamus. The direct projection of olfactory information to the neocortex sets olfaction apart from all other sensory systems, all of which enter the neocortex solely via the thalamus. The orbitofrontal cortex is involved in a variety of olfactory-related functions, including multimodal processing, judgments of edibility (is this a food-related odor?), and in higher cognitive functions such as naming and familiarity judgments.

Psychology

Since the first experimental psychologists studied olfaction in the late 1800s and early 1900s, a predominant focus has been on understanding the relation between the stimulus, the receptor that detects it, and the experience that it purportedly produces. Solving this so-called "stimulus problem" has been a major preoccupation of olfactory scientists. Such an emphasis is not surprising, as historically the same approach was adopted in audition (e.g., relation between sound frequency and pitch perception) and vision (e.g., relation between electromagnetic frequency and color perception). However, although research in audition and vision "moved on," with an increasing ecological focus in both on the perception of objects (e.g., the sound of "thunder" or the sight of a "tree"), olfactory research remained firmly anchored to a stimulus-problem approach.

There were benefits, but not ones researchers probably wanted. First, early investigators believed, reasonably enough, that there was only a limited set of receptors. It then followed that there should be only a limited set of primary odor qualities, with each primary associated with the activation of each specific receptor type (e.g., smoky, spicy, floral, etc.). That is, one should be able to identify a limited set of odor qualities (primaries), combinations of which would account for all possible olfactory experiences. No such set of odor primaries has ever been identified, and not for want of trying. Second, again, if there were a limited set of receptors, then some unlucky souls should be born without one or more receptor types (akin to color blindness) and this should allow one to determine the relation among stimulus, receptor, and percept. Again, researchers could not identify such consistent relations; those that they did identify differed significantly in form to their conceptual parallels in color blindness. Finally, although there are clearly relations between the percepts produced by chemically similar stimuli (i.e., aldehydes tend to smell fruity), it has never been possible to accurately predict what a novel chemical will smell like. For these reasons, although *empirical* research in the psychology of olfaction has made enormous strides over the last 100 years, *theoretically* there was little movement.

Things started to change in the 1980s with work on the neurophysiology of olfaction and with the genetic studies described above. The neurophysiology suggested that primary olfactory cortex functioned as a pattern-matching system, whereas the genetic data suggested what information this system might utilize. The implication of these developments slowly became apparent, especially when researchers took results from psychological work into account. A key study was conducted by Michael Rabin and William Cain (1984), who found that participants were poorer at discriminating unfamiliar odors and that this ability could be improved by passive exposure to such odors. This result, confirmed in a number of other laboratories, is very puzzling from a "stimulus-problem" perspective. The stimulus remains the same, but with increasing familiarity the person gets better at discriminating it from other unfamiliar odors. Moreover, Rabin and Cain then showed that the ability of people to discriminate the components of an odor mixture is also dependent upon familiarity. For example, if participants sniff an unfamiliar odor (A) and then they are presented with a mixture that contains this unfamiliar odor (A) mixed with *another* unfamiliar odor (B), they are very poor at saying whether A is present in the AB mixture. Performance improves markedly when one of the two odors is familiar and is best when both are familiar. Again, a stimulus-problem approach to olfaction is hard pressed to account for such findings, as the stimulus never changes.

The failure of the stimulus-problem approach to account for experimental findings has become more and more apparent. One such finding is that people can appear to smell things that are not in fact there. If participants are exposed to an unfamiliar odor mixture (e.g., smoky-cherry), when they later smell the cherry odor alone, they report that it smells slightly smoky. They also report, when smelling the smoky odor alone, that it smells slightly cherry-like. A more objective approach, using discrimination, reveals much the same result. If participants have experienced an odor mixture (e.g., smoky-cherry), they are much poorer at discriminating the two components (e.g., smoky odor vs. cherry odor) than they are at discriminating other odors that they have smelled an equal number of times, but never in a mixture. Findings of this type suggest a decoupling of what we perceive from what we are actually smelling and imply that what we are actually perceiving is a product of memory, not of the stimulus itself (i.e., in this case, a memory of smoky-cherry).

Findings such as these have led to a more contemporary theoretical view. When the olfactory system encounters chemical stimuli, it generates a pattern of neural activity that is then matched against previous patterns of activity that are stored in odor memory (this may reside in the primary olfactory cortex or it may be distributed around several structures, including the orbitofrontal cortex and parts of the limbic system). When there is a match between the incoming pattern and one stored in memory we perceive a particular odor, be it cheese or coffee. When there is no match, we experience a rather vague percept, one that is not readily discriminated from other such unfamiliar smells. In addition, when the pattern is unfamiliar it is also learned, such that when it is encountered again the odor starts to take on a more definite perceptual form—it becomes an odor object.

This type of mnemonic (memory-based) theory can readily account for many of the findings described above. For example, if one encodes an odor mixture (e.g., smoky-cherry), then later smelling the smoky odor alone should partially match the stored pattern (memory) of "smoky-cherry," resulting in perception of a cherry-like note in the smell. More important, it has no problem in accounting for why chemicals that share similar structural features (e.g., aldehydes) can sometimes smell alike, as structurally similar chemicals will produce similar patterns of neural activity and thus will activate similar odor memories. From this mnemonic perspective, the relation between the structural features of the chemical (the stimulus) and what we perceive is seen as correlational, but crucially not causal. The *cause* of a particular olfactory experience is the patterns that are activated in odor memory.

The strength of any particular theory is its ability to generate predictions and, as we alluded to above, to account for existing findings. We can derive two predictions from the mnemonic (memory) theory. First, as one encounters more odors, one should become progressively better at discriminating different smells. Put more bluntly, children should be poorer at odor discrimination than adults. Researchers have tested this claim extensively, using familiar, moderately familiar, and unfamiliar odors (as judged by adults). For all three odor types, adults and older children (11-year-olds) are consistently better at discrimination than six-year-old children. This difference could result from other causes. For example, adults and older children are better at naming odors and this also improves discriminative ability. However, using an articulatory suppression task (saying "the, the, the…."), which makes it hard to engage in any form of verbal processing such as naming, still reveals significantly poorer performance in 6-year-olds. Finally, it could be that younger children simply cannot perform the discrimination task. However, this possibility was also eliminated by having the children perform a complex visual discrimination task. Performance here did not differ from that of adults. Thus, younger children do indeed appear to be poorer at discriminating odors.

A second prediction is that any damage to odor memory should produce a profound deficit in olfactory perception, akin perhaps to visual agnosia, whereby a person can see (sensation) but cannot perceive visual objects (perception). There is a wealth of neuropsychological evidence in olfaction to support this type of deficit. The most striking example comes from the most well-known neuropsychological case, HM. HM had a bilateral temporal lobe resection, which included most of his primary olfactory cortex. The consequences for HM were profound, especially in the dense anterograde amnesia that followed his surgery (i.e., he could not remember anything new postsurgery). His olfactory perception was also profoundly affected. He could detect the presence or absence of odors as well as a normal person. He could also readily discriminate between a strong and a weak smell. However, he was totally unable to discriminate between different odors when they were equated for intensity (strength)—that is a rose, feces, and a fish supper all smelled alike. The conclusion that has been drawn from HM and similar cases is that even when the stimulus can be detected, if it cannot be recognized via the memory-based pattern-matching system, one cannot perceive odor quality—all odors come to smell alike.

Multisensory Processing

So far we have considered olfaction in a vacuum, yet, as noted above, when we smell we normally experience percepts in other sensory modalities at the same time. Olfaction is of special interest in this regard because of its involvement in "flavor." When we eat and drink a variety of independent sensory systems are activated, with smell, taste, and touch being the principal players. However, what we experience is a synthetic experience, especially between taste and smell. Paul Rozin reported that of 10 languages he surveyed, none had terms that clearly distinguished the olfactory component of eating and drinking, suggesting that at a linguistic level people are unaware of the contribution made by smell to flavor. Second, and more recently, several experiments have found that participants are very poor at discriminating the components of taste-smell mixtures in the mouth.

It is possible to account for these types of findings by recourse to much the same type of mnemonic model that we just discussed. For example, participants experience a particular combination of taste and smell, and these are then learned. Two consequences flow from this. First, the taste and the smell should now be harder to tell apart, in the mouth, than they were before. Researchers have not tested this prediction directly. Second, taste and smell might acquire each other's characteristics. Researchers have explored this prediction extensively. However, things are not as straightforward as one might first imagine. If, for example, you ask participants to "taste" sucrose (sweet) to which an odor has been added (say, lychee) and then later ask them to smell lychee odor alone, they generally report that the lychee odor now smells sweet, and, indeed, sweeter than it did prior to pairing with sucrose. However, if you now experience sucrose alone it does not result in an experience of lychee odor. This is a quite remarkable state of affairs, namely that an odor (a discrete sensory modality) can induce perception of a taste (another discrete sensory modality)—synesthesia. More recent evidence suggests that animals too may experience something substantially similar.

Critics have argued that the perception of sweetness produced by an odor is metaphoric—we speak of "sweet" people but we do not mean that they *taste* sweet. The research evidence does not support this interpretation. Odor sweetness is akin to that produced by sucrose on the tongue, and this result is suggested by the phenomenon of sweetness enhancement. Here a tasteless but sweet-smelling odor is added to a sucrose solution and this results in participants rating this mixture as sweeter than the sucrose alone. The degree to which this enhancement occurs is best predicted by how sweet the odor is judged to smell, suggesting that smelled "sweetness" and tasted sweetness are perceptually similar.

The experience of odor sweetness is almost certainly mnemonic. Nothing about the odor *stimulus* changes as a consequence of experiencing an odor and sucrose together. The similarity to the results described earlier lends yet more weight to the notion that what we experience when we smell an odor is not primarily driven by what is actually there, but rather by what it reminds us of—redolence.

METHODS

Experimental approaches to studying olfaction in both humans and animals utilize many methods common to other areas of sensory psychology—threshold, discrimination, identification, and so on. However, olfaction offers certain specific challenges along with some rather specialized methods; these are examined in the following.

The Stimulus

The selection of odors for an experiment requires careful attention. As should now be apparent, individual experience with an odor can affect its discriminability. We also know that individual experience with an odor tends to produce (in the main) more positive hedonic responses toward it and that such odors are also judged to smell stronger. A more general issue concerns the use of single pure chemicals or more ecologically valid stimuli (e.g., coffee) in experiments. Although single pure chemicals offer ease of replication, they do not reflect the type of stimulus that the olfactory system generally has to contend with. Certain researchers, such as Robyn Hudson (1999), have argued that progress in olfactory research has been slowed by an unwillingness to utilize complex ecologically valid odors.

When selecting odors, care should be taken to specify the experimental requirements and the population under test. For learning experiments, odors probably need to be unfamiliar to most participants, although for identification, the odors likely need to be familiar. With older participants, who are known to have poorer odor naming and acuity and reduced perception of odor intensity, more intense stimuli might be needed to compensate for such deficits. In younger participants, especially children, care needs to be taken with toxicity and with the presence of potential allergens, especially products containing nuts.

Adaptation

A special problem faced by olfactory researchers is in providing sufficient time between stimuli to prevent adaptation from occurring. Unfortunately, there are no well-standardized guidelines upon which to base a decision, although E. P. Kosters's doctoral thesis on odor adaptation probably has the most complete set of data about adaptation;

it is a valuable source of information. Whereas a visual object remains visible throughout the viewing period, repeated presentation of an odorant will lead to progressive loss of the ability to perceive it with possible cross-adaptation to other odors, too. This varies on both an individual basis and on an odor-by-odor basis, making it hard to prescribe general guidelines. Nonetheless, at least 15 to 20 seconds should intervene between each stimulus, with longer intervals (60 seconds) if the same stimulus is used repeatedly.

Specialized Methods

Two specialized methods warrant mention. The first is the most widely used neuropsychological test of olfactory ability, the Smell Identification Test developed by Richard Doty (SIT, previously known as the UPSIT; 1997). This test is composed of 40 microencapsulated odors. For each odor, the participant scratches the microencapsulated dot and releases the smell, which the person then sniffs. The person then has to identify the odor by selecting one of four response options provided. The test is scored out of 40 and norms are available by age and gender. This test has been used extensively to explore olfactory deficits in neuropsychological and psychiatric populations, in industrial settings, as a screening test, and with normal participants in a variety of experimental contexts. The test has excellent reliability and validity and provides a snapshot of a person's olfactory capabilities. Its major limitation is that it does not indicate the precise nature of any deficit.

For certain types of olfactory research in both animals and people, precise metered delivery of an odorant is required. This typically involves a specialized piece of equipment—an olfactometer. These are usually built to order, and vary in the speed and accuracy with which they can deliver an odorant to the nose and in their ability to do so without any auditory cue (i.e., clicking of valves). The number of channels is also an important consideration, in terms of the number of odors the researcher can readily administer within an experimental session. Delivery may also need to be to breathing rate, so that the odor is delivered at the moment of inspiration. Several research groups around the world build olfactometers, and published details are available in the References about how to construct one.

APPLICATIONS

The psychological study of olfaction has applications to many diverse areas; I examine a few important examples in this section. The place to start, perhaps, is perceptual expertise, as this relates back in many ways to issues raised early about learning and memory. Within the olfactory domain perceptual expertise manifests in many forms: wine, beer, dairy, and cheese tasters; flavorists (who create food flavors); perfumists (who create perfumes and many perfume-like additives for laundry powder, wood polish, etc.); and expert sensory evaluation panels. All of these individuals usually have a fairly extensive training that typically involves learning the name for many of the chemical compounds that they may encounter within their domain of expertise.

There have been several investigations of how perceptual experts differ from novices. One apparently surprising finding is that such individuals are not as good as one might at first believe. Claims that wine tasters may identify the soil type, hillside, vintage, and so forth are exaggerated. Similarly, the often verbose descriptions of a wine's qualities are unreliable, as research has shown that when experts are asked to provide such written descriptions and then later match them to the same set of wines, although they are above chance and significantly better than naïve controls, the actual size of the effect is small. Moreover, when assessed purely in their discriminative capacity (i.e., to tell one wine from another), experts fare no better than *amateur* wine drinkers. Their expertise only becomes apparent when they are asked to provide a verbal description. It appears that experts may move back and forth between a perceptual experience (the wine's flavor) and a verbal description more readily than amateur drinkers (i.e., experts are less prone to verbal overshadowing).

One way to eliminate the need for experts, and perhaps sniffer dogs as well, is to design electronic noses that may detect tainted food, drugs, explosives, contraband food, or whatever. Although there has been some success in this regard, nothing can yet match the acute sensitivity and accuracy of a dog's nose. Moreover, dogs can be trained to identify several different targets, whereas most electronic noses are limited to detecting a single chemical and may have trouble with odors in which there are considerable variations in chemical composition. In fact, it is likely that major advances in electronic noses will come from software that mimics the type of associative memory with which we have been endowed.

Some smells may genuinely indicate danger (e.g., smoke, natural or poison gas). In the United States alone, 3 million people live near an oil refinery, with exposure to high levels of aromatic hydrocarbons, and over 33 percent of Americans regard air pollution as a major problem. Certain environmental odors can be offensive and produce significant loss of enjoyment for those forced to live with them, even though the level of the pollutant may not be sufficient to directly cause ill health. These problems may occur from traffic fumes, paper mills, chemical plants, sewage farms, and animal husbandry (abattoirs, feed lots, etc.). However, these types of odors may be *perceived* as noxious and injurious to health and produce somatic symptoms, consistent with the person's beliefs about the odor. Tobacco smoke is a case in point. Whereas 30 years ago few complained about having to share public space with smokers and their smoke, changing perceptions of what secondhand smoke *means* have led to a "moralization" of tobacco smoke. This allows people to express outrage at smokers for polluting the local environment, even when limited exposure can produce no detectable harm. More generally, the precise measurement of noxious odors is problematic, as they rarely correlate with subjective impressions of the problem. Nonetheless, there is a developing research agenda that seeks to understand how people's beliefs affect their perception of odors and how exposure to noxious smells influences health and behavior.

Tests of olfactory function may provide useful warning signs of impending disease. In people who have a heightened risk of developing schizophrenia, poor odor naming is the single best predictor of whether they will later develop this disease. Similarly, the earliest indicator of the presence of Alzheimer's disease is a notable reduction in odor-naming ability (and discriminative capacity), which may occur many years before more overt cognitive deterioration is evident. The use of odor naming as a prognostic tool is still in its earliest stages, but it is likely to play an important role in the future, especially as tests such as the SIT (see Specialized Methods section) appear very sensitive to abnormalities in frontal-lobe function (e.g., in ADHD, OCD).

Finally, the study of olfaction in animals has a number of practical applications. The identification of various insect pheromones has led to the development of insect traps that utilize pheromones to attract pest species. Similarly, with mosquitoes, there have been significant advances in understanding what makes some individuals more prone to be bitten than others, and it is likely that new forms of chemical control will eventuate from this research. In addition to new weapons against insect pests, advances in rodent control based upon their innate fear response to odors that are secreted by cats and foxes are also under development. These odors can be readily collected from cats and foxes, but researchers have yet to identify the key ingredients.

COMPARISONS

Four forms of comparison, which have all attracted considerable research attention, appear to be particularly valuable:

1. Microsmatic animals (sense of smell is of lesser importance; e.g., humans and primates) versus macrosmatic animals (sense of smell is very important; e.g., rodents and dogs)
2. The main olfactory system (on which this chapter is mainly focused) versus the accessory olfactory system (a system that functions to detect *specific* chemicals, often related to reproduction)
3. Men versus women
4. Olfaction versus other sensory modalities in humans

Microsmatic Versus Macrosmatic Mammals

Animals differ in their sensory specialization—well-developed eyes are of little use to night-flying bats or rodents. For olfaction, there has been a traditional divide between microsmatic and macrosmatic animals. Although this divide has some utility, it probably reflects our ignorance about the olfactory capabilities of many species that have never been examined in any depth (e.g., monotremes, whales, dolphins). Nonetheless, some mammals clearly have a greater reliance on their sense of smell, notably dogs, rodents, and certain carnivores. In these animals, especially dogs and rodents, which have received the most attention, the size of the olfactory mucosa is proportionally larger and contains more olfactory receptor neurons than that in species with a lesser reliance upon smell such as humans. Whereas we have around $10cm^2$ of olfactory epithelium, dogs such as German shepherds have up to $169cm^2$, boxers have up to $120cm^2$, and even Pekinese have around $30cm^2$. Similarly, the number of olfactory receptor neurons is also greater in dogs—humans have around 10 million, fox terriers have 150 million, and sheepdogs have 220 million. Functionally, this allows animals such as dogs to detect odors at concentrations well below what we can detect. Consequently they are sensitive to trace levels of drugs or explosives, or even underground gas leaks. However, it is not clear what other benefits this sensitivity may confer. As we described, much olfactory processing relies upon higher cortical functions (recognition), and although rodents and dogs may have proportionally more cortex devoted to olfaction than humans, the degree to which they are better at recognition may not be so large. Clearly there are some advantages. Dogs can, for example, readily discriminate the odors of different people (which we find hard) and even discriminate identical twins (with difficulty), whereas rats appear able to analyze odor mixtures in a way we cannot. Their main advantage, however, is sheer sensitivity.

Stimulus-Driven Versus Mnemonic Olfactory Systems

So far we have concentrated on general olfactory abilities and thus on the main olfactory system. However, nearly all animals, both vertebrates and invertebrates, have both a main and an accessory olfactory system. Whereas the main olfactory system is plastic and flexible—that is, it can recognize many odors and acquire new ones—the accessory olfactory system is characterized by its general inflexibility and specificity. That is, it appears to be hardwired to detect certain specific chemicals or mixtures of chemicals. Not surprisingly, these turn out to be chemicals that often have important biological functions, and many of these fall under the rubric of "pheromones"—chemicals secreted by one member of a species that have an instinctual effect on the behavior or physiology of another member of that species.

Whereas the existence of an accessory olfactory system and pheromones in insects, fish, rodents, and some other mammalian species as well (e.g., rodents) is solidly established, the existence of both an accessory olfactory system in humans *and* human pheromones is highly controversial. There clearly is evidence that humans have cells located in the nasal septum that *appear* similar to those found in other mammalian accessory olfactory systems (the vomeronasal organ), but whether these cells can actually detect chemicals and influence behavior has not been established. As for the effect of pheromones on human behavior, this is a field awash with claims, some of which are supported and many of which rely on weak evidence. Menstrual synchrony driven by chemical cues is probably the most replicated finding, albeit not without its critics. Here, women who live in close proximity tend to progressively synchronize their menstrual cycles. However, whether or not certain chemicals such as 5-alpha-andrsotenone act as a chemical attractant (secreted by men and attracting women) is a highly controversial claim and is currently unresolved, although such chemicals are widely marketed.

Men Versus Women

On virtually every sort of olfactory task—sensitivity, recognition, and identification—women tend to outperform men. This finding is well established and robust, although its cause is not well understood. One possibility is that higher circulating levels of estrogen in women may enhance olfactory function. This might explain why olfactory abilities are better preserved into old age in women than in men, as well as the variations in olfactory acuity that occur in women during the menstrual cycle. It does not explain the observation, however, that prepubescent females may exceed prepubescent males on olfactory tasks, although the data here are somewhat limited and not all studies report such differences.

A second possibility is that women may simply attend to olfactory stimuli more than men—they certainly dream of smells more frequently—as they may have a greater interest in things olfactory as a consequence of gender stereotyping (i.e., cooking, perfume, etc.). Recent attempts to test this reveal that women still outperform men even when unfamiliar odors are used (i.e., those with which they could have had no prior experience) and where level of motivation does not differ either. Moreover, *National Geographic* conducted a large smell survey of over 1 million people during the early 1990s, across many different nations and cultures (i.e., with presumably different types of gender stereotyping), and yet this too found consistent female advantage in performance. Finally, some have suggested that women's better performance on olfactory tasks is an evolutionary adaptation reflecting their putative role in gathering food supplies of plant origin (whereas men hunted). It has been argued that this would place a selective pressure on women for keener olfactory capabilities, so as to distinguish edible plants and avoid toxic ones. Such evolutionary explanations may account for superior female abilities, but they do not explain their biological basis.

Olfaction Versus the Other Senses

At both the neuroanatomical and the psychological level, olfaction differs qualitatively from the primary senses of vision and audition. As we noted earlier, olfaction has a unique neuroanatomy in being the only sense to have both a direct route (to the orbitofrontal cortex) and an indirect route (via the mediodorsal nucleus of the thalamus) to the neocortex. The functional significance of this distinction is a matter of debate, but some have suggested that it might account for the universal form of synesthesia (i.e., "tasty smells") that olfaction exhibits.

At the psychological level, olfaction appears to differ in several respects. The first concerns the presence of discrete short-term odor memory capacity, as distinct from a long-term odor memory capacity. Such a distinction is usually made on four grounds—capacity differences, coding differences, neuropsychological dissociations, and the serial position effect. To date, no consistent evidence from any of these domains has been identified, and although some contend that olfaction does have a discrete short-term store, this is contentious.

A second difference is that olfaction appears to be primarily a synthetic sense with little available capacity to discriminate the component parts of the stimulus. In contrast, although we may appreciate that a telephone is one visual object or that "Penny Lane" is a Beatles song, we can identify many individual components within these stimuli, a feat that we cannot readily accomplish in olfaction even after specialist training.

A third difference is the more constrained capacity for odor imagery. Some suggest here a qualitative difference, claiming that olfaction has no imagery capacity whatsoever, and there is certainly evidence to back this claim. Even where evidence of imagery is obtained it tends to be more circumscribed in its capacity than in vision or audition, but the evidence of olfactory hallucinations strongly suggests that under certain circumstances people clearly do perceive an odor when one is not in fact there.

A fourth difference, and the one that has been most extensively explored, is the olfactory-verbal "gap." Most people can readily name visual objects, yet most people have enormous difficulty naming common odors when their visual cues are absent (e.g., sniffing coffee from an opaque bottle). The origins of this effect are poorly understood, even though the phenomenon itself is robust. Tyler Lorig has suggested that olfaction and language may compete for the same neural resource, making it difficult to name an odor. Others suggest that it is the neuroanatomical remoteness of language areas from primary olfactory processing areas and the sparsity of connections between them that result in impoverished naming. More prosaically, it may in part be a consequence of practice, as teaching participants to name odors does produce a rapid improvement in performance, and outside of the laboratory most of us do not have to rely upon the blind naming of odors.

Finally, it has long been claimed that odors may serve as powerful retrieval cues for distant autobiographical memories—the Proustian phenomenon (Proust describes how the smell of a madeleine biscuit dipped in tea led to the evocation of forgotten childhood memories). Certainly, odors do appear able to act as retrieval cues for older memories in adult participants, more so than comparable nonolfactory cues, and odors clearly invoke memories that are more emotional as well. But as with so many things olfactory, what we do not know far exceeds what we do.

FUTURE DIRECTIONS

For nearly all of the areas discussed in this survey of olfaction, the paucity of data is striking. Olfaction is heavily underresearched, a likely consequence of our reliance as a species on vision and audition. However, the study of olfaction is important for reasons that may not be readily apparent. First, it offers a glimpse into how the visual and auditory object recognition systems may have developed. This is because a developed chemical sense predates, in evolutionary terms, complex visual and auditory systems. Second, the olfactory system is an ideal place to study consciousness. In relative terms, the system is far less complex than the visual and auditory systems and so may provide a better place to tackle the hard question of consciousness (i.e., why smell *feels* different from sight or hearing and, more generally, how we *experience* such feelings at all). To give a more specific example, the apparently universal synesthesia that is demonstrated in the olfactory system ("tasty smells") offers an excellent tool with which to explore the nature of consciousness. Since the time of William James, scientists have pondered whether specific areas of the neocortex were associated with specific sensory qualities (e.g., visual or auditory, etc.). Synesthesia is an excellent way of studying this problem—that is, if a person can no longer *taste* sweet things, can he or she still experience sweet smells? Do these similar percepts rely upon the same area of the brain? Third, whereas olfaction may be less important to humans, it is certainly very important to many other species—and determining how the system functions may lead to important advances in the control of agricultural pests, insect vectors of disease (e.g., mosquitoes, and pest animals such as rodents.

Perhaps the most important area of research activity in olfactory psychology at the moment is in determining whether (and how) odor object recognition may function. As of the writing of this chapter, there is a growing consensus, both empirically and theoretically, that this is a more useful way of thinking about olfaction instead of concentrating on the relation between the stimulus (odor) and the response it produces. There are as yet almost no research papers that actually set out to test predictions derived from odor object recognition theory, indeed the theory itself is still poorly developed. It is these sorts of issues that will likely capture the attention of a new generation of olfactory psychologists in the 21st century.

REFERENCES AND FURTHER READINGS

Baxi, K. N., Dorries, K. M., & Eisthen, H. L. (2006). Is the vomeronasal system really specialized for detecting pheromones? *Trends in Neurosciences, 29*, 1–7.

Brand, G., & Millot, J.-L. (2001). Sex differences in human olfaction: Between evidence and enigma. *Quarterly Journal of Experimental Psychology, 54B*, 259–270.

Brewer, W. J., Wood, S. J., Phillips, L. J., Francey, S. M., Pantelis, C., Yung, A. R., et al. (2006). Generalized and specific cognitive performance in clinical high-risk cohorts: A review highlighting potential vulnerability markers for psychosis. *Schizophrenia Bulletin, 32*, 538–555.

Buck, L. B. (1996). Information coding in the vertebrate olfactory system. *Annual Review of Neuroscience, 19*, 517–544.

Dielenberg, R. A., & McGregor, I. A. (2001). Defensive behavior in rats towards predatory odors: A review. *Neuroscience and Biobehavioural Reviews, 25,* 597–609.

Doty, R. L. (1997). Studies of human olfaction from the University of Pennsylvania smell and taste center. *Chemical Senses, 22*, 565–586.

Eichenbaum, H., Morton, T., Potter, H., & Corkin, S. (1983). Selective deficits in case H. M. *Brain, 106*, 459–472.

Freeman, W. J. (1983). The physiological basis of mental images. *Biological Psychiatry, 18,* 1107–1125.

Haberly, L. B. (2001). Parallel-distributed processing in olfactory cortex: New insights from morphological and physiological analysis of neuronal circuitry. *Chemical Senses, 26*, 551–576.

Herz, R. S., & Inzlicht, M. (2002). Sex differences in response to physical and social factors involved in human mate selection: The importance of smell for women. *Evolution and Human Behavior, 23*, 359–364.

Hudson, R. (1999). From molecule to mind: The role of experience in shaping olfactory function. *Journal of Comparative Physiology, 185A*, 297–304.

Hughson, A. L., & Boakes, R. A. (2001). Perceptual and cognitive aspects of wine experts. *Australian Journal of Psychology, 53,* 103–108.

Jacob, S., & McClintock, M. K. (2000). Psychological state and mood effects of steroidal chemosignals in women and men. *Hormones and Behavior, 37*, 57–78.

Jonsson, F. U., Tchekhova, A., Lonner, P., & Olsson, M. J. (2005). A metamemory perspective on odor naming and identification. *Chemical Senses, 30,* 353–365.

Livermore, A., & Laing, D. G. (1996). Influence of training and experience on the perception of multicomponent odor mixtures. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 267–277.

Lorig, T. S., Elmes, D. G., Zald, D. H., & Pardo, J. V. (1999). A computer-controlled olfactometer for fMRI and electrophysiological studies of olfaction. *Behavior Research Methods, Instruments & Computers, 31*, 370–375.

Mainland, J., & Sobel, N. (2006). The sniff is part of the olfactory percept. *Chemical Senses, 31,* 181–196.

Morot, G., Brochet, F., & Dubourdieu, D. (2001). The color of odors. *Brain and Language, 79*, 309–320.

Murphy, C. (2002). Olfactory functional testing: Sensitivity and specificity for Alzheimer's disease. *Drug Development Research, 56,* 123–131.

Porter, J., Craven, B., Khan, R., Chang, S., Kang, I., Judkewicz, B., et al. (2006). Mechanisms of scent-tracking in humans. *Nature Neuroscience*, 10, 27–29.

Rabin, M., & Cain, W. S. (1984). Attention and learning in the perception of odor mixtures. In D. G. Laing, W. S. Cain, R. L. McBride, & B. W. Ache (Eds.), *Perception of complex smells and tastes* (pp. 173–188). Sydney: Academic Press.

Raman, B., Sun, P. A., Gutierrez-Galvez, A., & Gutierrez-Osuna, R. (2006). Processing of chemical sensor arrays with a biologically inspired model of olfactory coding. *IEEE Transactions on Neural Networks, 17*, 1015–1024.

Rozin, P. (1982). "Taste-smell confusions" and the duality of the olfactory sense. *Perception and Psychophysics, 31*, 397–401.

Shepherd, G. M. (2005). Outline of a theory of olfactory processing and its relevance to humans. *Chemical Senses, 30*, 3–5.

Shepherd, G. M. (2006). Smell images and the flavour system in the human brain. *Nature, 444,* 316–321.

Stevenson, R. J., & Boakes, R. A. (2003). A mnemonic theory of odor perception. *Psychological Review, 110*, 340–364.

Verhagen, J. V., & Engelen, L. (2006). The neurocognitive bases of human multimodal food perception: Sensory integration. *Neuroscience and Biobehavioural Reviews, 30*, 613–650.

Wilson, D. A., & Stevenson, R. J. (2003). The fundamental role of memory in olfactory perception. *Trends in Neuroscience, 26*, 243–247.

Wilson, D. A., & Stevenson, R. J. (2006). *Learning to smell: Olfactory perception from neurobiology to behavior.* Baltimore: Johns Hopkins University Press.

25

AUDITION

MICHAEL FIRMENT
Kennesaw State University

We use sound to communicate, to identify and locate sound-emitting objects, and to provide us with entertainment through the making of and listening to music. The study of these functions makes up large portions of the psychological study of audition under the categories of speech perception, auditory localization, scene analysis, and music perception. Understanding these topics often requires knowledge of psychological areas that may at first seem to have little relation with them but that are of great importance for understanding them. Some knowledge of the physics of the sound stimulus and of the anatomy and physiology of the ear and brain is needed for the fruitful study of auditory perception. Related fields include *acoustics* in engineering and the biological and medical study of hearing.

Knowledge of the physics of sound and the structure and function of the ear and brain can be important for the design and interpretation of experiments that may seem purely psychological. For example, early estimates of the frequency range of human hearing ranged from lows of 8 Hz to highs of 55,000 Hz, ranges that are considerably different from the currently accepted 20 Hz to 20,000 Hz for young adults (Hertz, Hz, and cycles per second, cps, are synonymous). The errors were mostly caused by a lack of consideration for the overtones (frequencies higher than the fundamental frequency, called harmonics when they are multiples of the fundamental frequency) produced by the whistles and tuning forks used in the studies (Boring, 1942).

Although psychologists have been and continue to be involved in anatomical and physiological investigations of the ear and the brain, the psychological study of hearing is primarily based on human behavioral responses to sound stimuli. The measures include simple yes-no responses, the estimation of magnitude, ratings, interpretations of speech sounds, and many others. A good knowledge of psychophysical techniques and their limitations is important (see Krantz's Chapter 20 on Psychophysics).

Experimental work in audition is motivated by both applied and theoretical concerns. Applied work might include topics in human factors such as how changes in hearing might affect the elderly when they drive, or how blind people might use ambient sound to orient themselves in an environment. Auditory research can also be applied to clinical areas, such as how hearing loss might increase paranoid thoughts (Zimbardo, Anderson, & Kabat, 1981). More theoretical investigators may concern themselves with the physiology of the ear and brain or with more traditional psychological topics. For example, Cusack and Carlyon (2003) looked into similarities between the feature integration theory of visual object perception and the discrimination of auditory features such as sound duration.

HISTORY

Researchers have studied sound, hearing, and music for thousands of years. Pythagoras (approximately 582–510 BCE) used plucked strings to note the relation between the length of the string, the pitch of the note, and the musical quality of the note. He (and, no doubt, many others) noted that when the string's length is decreased to one half of its original length, the higher pitch produced was that of the same note one octave above the original note (Boring, 1942). As we now know, the string's rate of vibration is proportional to its length, and an increase of one octave is a doubling of the original frequency.

Aristotle (approximately 384–322 BCE) wrote that sound was produced when air was moved in a sudden, powerful way so that it moved as a single packet and did not dissipate. He also noted that sound could be heard in, and therefore carried by, both air and water. He was correct in that air has to be moved quickly in order to produce audible sound, but he did not describe the continuous vibrations present in most sounds.

The scientific study of sound and hearing began when craftsmen and physicists produced instruments to measure and produce sounds. Their early experiments used tuning forks and sirens to produce pitches, and they obtained variations in loudness by such devices as varying the weights of steel balls dropped on ebony plates (Boring, 1942). Precise and relatively inexpensive control of intensity and frequency became possible when electronic amplifiers and oscillators were invented in the early 1900s. Also, as an indication of the greater precision made possible by electronics, pitches began to be given in terms of frequency rather than in terms of musical notes.

Physicists such as Helmholtz, Mach, and Ohm did early work on sound and hearing. Ohm's acoustic law stated that we are able to perceive the separate sine wave components of a complex tone that would be revealed when doing Fourier analysis of the tone. Ohm's law complemented Helmholtz's theory that the basilar membrane of the inner ear was made up of a series of resonators, each most sensitive to a different sound frequency. The physicist Max Wien made an early measurement of the lower limit of sound intensity that is audible. Smith calculated that the amount of energy that a blade of grass used as it grew and raised itself from the ground, if translated into a sound frequency of about 4,000 Hz, would be perceived as the same loudness as a quiet bird call or the sound of a refrigerator running—that is, about 40 dB (Jones, 1937).

Early psychologists such as Stumpf, Munsterberg, and later Stevens did much work in clarifying how—and how well—people could locate sound-emitting objects (Boring, 1942). Looking through the early volumes of the *Journal of Experimental Psychology*, one can find articles on whether pitch and volume have separate psychological realities (Rich, 1916) and on the possibility of producing illusions of movement using sound stimuli (Burtt, 1917). Both of these articles, done prior to the widespread availability of electronic sound-generating equipment, devoted considerable space to describing their custom-built experimental apparatus.

PHYSICAL STIMULUS

Sound is created by a rapid change in pressure in the atmosphere (or some other material) over time, caused by a vibrating solid (or some other source) that alternately compresses and rarifies (decompresses) the air next to it. These changes are rapidly communicated to the neighboring air molecules, creating compression waves, or frequencies. We can perceive, depending on our age and the intensity of the sound, frequencies between about 20 and 20,000 Hz (cps). These frequencies produce extremely small movements of our eardrum that lead to movements of the hair cells (cilia) in the inner ear, which are then translated into nervous impulses to be processed in the brain. The intensity—the amount of pressure change—is generally given in terms of decibels, $dB = 20 \log_{10} (p_1/p_0)$. This is a compressed scale that allows a very large range of pressure differences to be described within a reasonable range. Pressure ratios from 1 to 10,000,000 are given by a range from 0 to 140 dB. The p_0 in the equation is generally set to 20 microPascals, a value based on the human absolute threshold for hearing a 1,000 Hz tone. When calculated in this fashion, the dB level is called a *sound pressure level* (SPL). Occasionally p_0 is set to the participant's personal threshold, and the value is known as the *sensation level*. Generally, the greater the pressure change, the louder the sound.

The number of alternations between high and low pressure per unit time—the *frequency*—is given in Hertz (Hz), or cycles per second. Generally, the higher the frequency, the higher the perceived pitch. The pitch of a complex tone, anything other than a simple sine wave-shaped tone, depends on factors other than the average frequency. For example, if sine wave tones of 600, 800, 1,000, and 1,200 Hz were combined, the perceived pitch would be similar to one produced by a 200 Hz tone.

The most natural depiction of sound is in a figure that shows the intensity on the *y*-axis and time on the *x*-axis. This can be called a description of the sound on the *time domain*. Because waveforms can be synthesized by combinations of simple sine waves of varying frequencies, intensities, and phases, sound can also be described on the *frequency domain* with intensity again on the *y*-axis and frequency on the *x*-axis. (Figures 25.1 and 25.2 show a combination of a 300 and a 500 Hz tone in the time and frequency domains.) It is also possible to combine these two domains with a spectrogram. In this display, time is on the *x*-axis, frequency is on the *y*-axis, and intensity is encoded by color (Figure 25.3).

Sound waves, like almost any other sort of wave, can be broken down into a set of sine waves (a wave that changes in magnitude in ways that can be mapped onto the

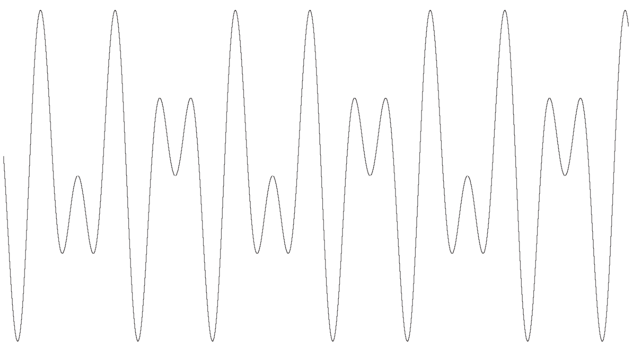

Figure 25.1 A combined 300- and 500-Hz sine wave tone in the time domain.

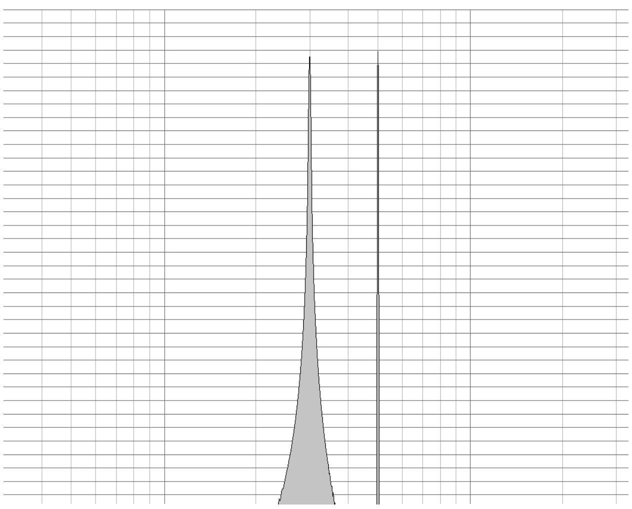

Figure 25.2 A combined 300- and 500-Hz sine wave tone in the frequency domain.

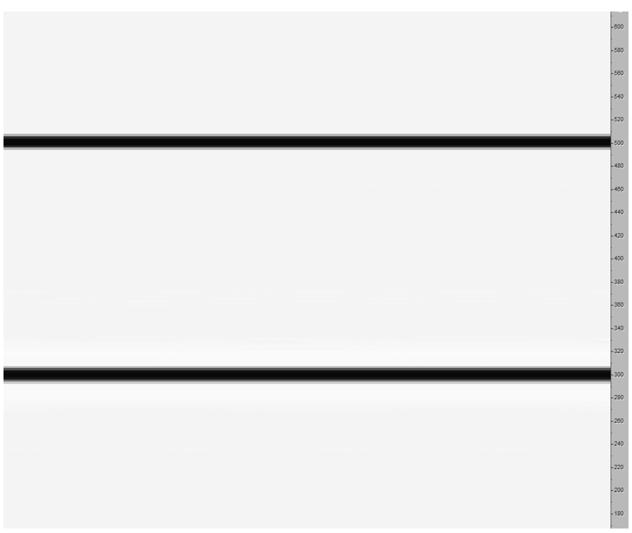

Figure 25.3 A combined 300- and 500-Hz sine wave tone in a spectrogram display.

Figure 25.4 Overtones of a flute.

trigonometric sine function) using a method discovered by Fourier (1768–1830). It appears that the ear does something similar to a Fourier analysis of sound waves, breaking them down into a set of simple sine waves.

When our voices or musical instruments produce a tone—a particular frequency—they also produce tones at integer multiples of that frequency. These overtones, or harmonics (or partials), are part of what goes into the particular sound or *timbre* of the voice or instrument. (Figure 25.4 shows a spectrogram of the fundamental frequency and harmonics produced by a flute.)

It is often important to use higher-order descriptions such as how the intensity of a sound varies with time (amplitude modulation) or how its frequencies change with time (frequency modulation). For example, the difference between a piano note and an organ note (the difference in timbre) is due, in large part, to the way the intensity of the note changes with time, its *intensity envelope* (Figure 25.5). And, the identification of words requires noting the changes in frequencies of vowel *formants*, frequencies in the speech signal having relatively high intensities (Figure 25.6).

The sounds we hear are made up of noise and tones. Although we can define noise as any unwanted sound, in this discussion it refers to sound that is made up of a more or less random collection of frequencies that each exist for short periods of time. A tone is a continuous (although possibly changing) frequency. Our voices are made up of a combination of the two. Very roughly, vowels are made up of changing tones and consonants are made up of noises (often accompanied by tones). There are several named types of noises. *White noise*, which sounds something like two dry hands being rubbed together, is made up of frequencies that are all of the same average energy. Due to the way our hearing system works, it sounds rather high-pitched. *Pink noise* has equal energy per octave—that is, the interval of frequencies from 200 to 400 Hz has the same amount of energy as the interval between 400 and

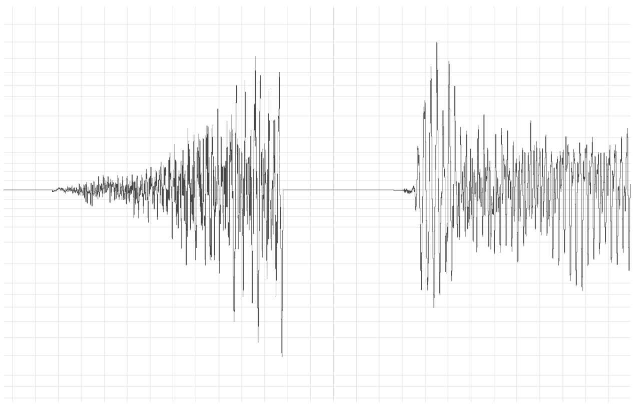

Figure 25.5 Tone envelope of an organ (to the right) and a piano.

Figure 25.6 Spectrogram of a voice showing formants.

800 Hz or between 800 and 1,600 Hz. Both pink and white noise can be used to mask background noise or the perceived ringing noise produced by tinnitus. Pink noise is particularly useful in setting up audio equipment through the use of analyzers and equalizers that are calibrated in octaves or fractions of octaves. Another category of noise, *band-pass noise*, noise from which upper and lower frequencies have been eliminated, is often used in auditory experiments.

THE EAR AND THE BRAIN

The structures that make up the ear transform the sounds presented to us by the outside world into neural signals. This output mirrors the time relations in the original sound and encodes the frequencies in the sound through a combination of duplicating the frequency of the sound (temporal coding) and carrying the information concerning different frequencies with different neurons (place coding).

When the sound acts to vibrate the eardrum (tympanic membrane), it has already been modified by the shape of the side of the head, the external ear (pinna, auditory meatus), and the ear canal. The contours of the outside structures both filter and amplify certain frequencies, the pattern of the modification depending a bit on the direction the sound is coming from. This directionally dependent filtering can help us locate auditory objects. The shape of the entrance to the ear canal and the ear canal produces a horn that amplifies frequencies between about 1,000 and 4,000 Hz.

Water is harder to move than air. Sound that travels freely through the air will be simply reflected by water. This can be seen from the fact that noises coming from the outside world are only faintly heard by someone under water. In order to affect the sensory cells in the fluid-filled inner ear, the motions of the air molecules must be concentrated, decreased in extent but increased in pressure. The structures of the middle ear—the eardrum and ossicles (three small bones, the malleus, incus, and stapes)—serve this purpose, matching the impedance (ease of movement) of the air with that of the fluid in the cochlea. The eardrum has an area of about 55 mm^2. After passing through the ossicles, the smallest bones in the body, the size of the end plate of the last ossicle, the stapes, is about 3.2 mm^2. This difference in size increases the pressure by a factor of 17 (Luce, 1993). This increase, along with leverage produced by the ossicles, results in an increase of pressure by about 30 times.

The ossicles are attached to two muscles that can decrease their mobility. Intense sounds cause these muscles to contract and lessen the amount of sound that reaches the inner ear, especially those of lower frequency. The acoustic reflex probably has several functions. It allows very loud sounds, sounds that would normally be too powerful to be processed by the hair cells, to be perceived more accurately. Another plausible function is to prevent the low-frequency speech sounds that are transmitted to the middle ear through the bones of the skull from interfering with our perception of outside events (Buser & Imbert, 1992).

The inner ear contains a complex active mechanism that separates in a very precise way the different frequencies that make up a sound into different positions along the length of the inner ear. The cochlea (most typically pronounced with an *ah* sound rather than an *oh* sound), where sound is transduced into nervous impulses, is a hollow area shaped somewhat like a snail's shell. If it could be straightened its length would be about 3.5 cm. It is divided lengthwise by two membranes. Reisner's membrane separates the upper chamber, the scala vestibuli (vestibular canal) from the scala media (cochlear canal). The bottom of the scala media is made up of the basilar membrane upon which the sensory cells for audition, the hair cells, rest. The area below the basilar membrane is the scala tympani (tympanic canal). At the apex of the cochlea (the tip, if it were straightened) is a small hole, the helicotrema, that joins the scala vestibuli and scala typani. The stapes is connected to the scala vestibuli by means of the oval window, which is a membrane-covered oval hole in the vestibule of the inner ear. In order to allow the stapes to move the fluids in the cochlea, a second flexible membrane-covered hole, the round window,

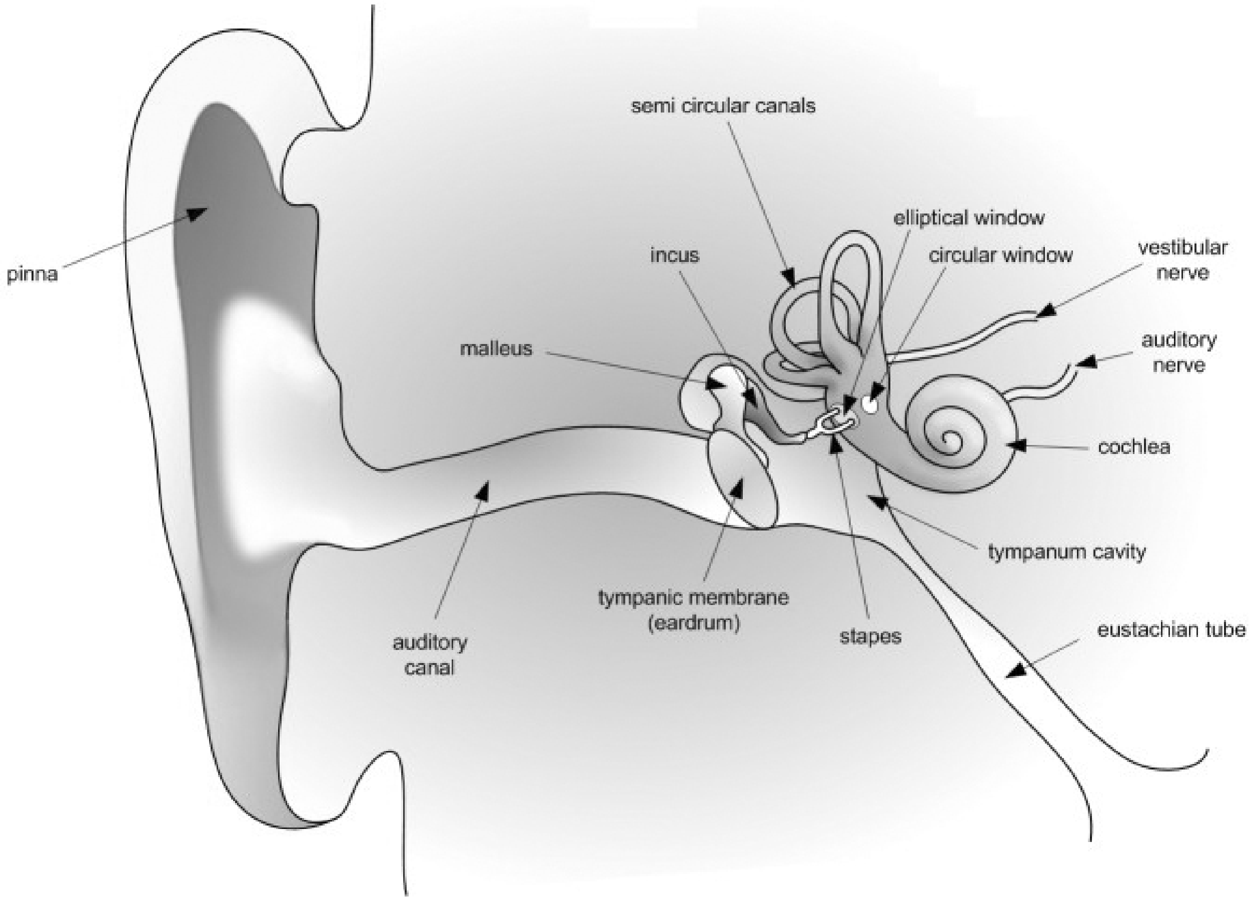

Figure 25.7 The ear.

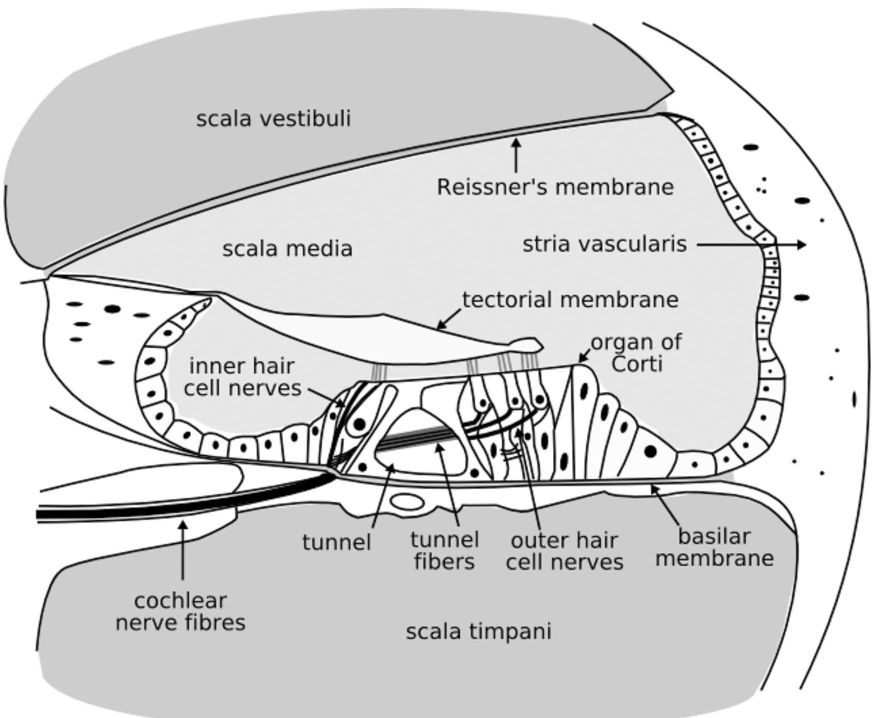

Figure 25.8 Cross-section of the cochlea.

is located at the base of the scala tympani.

The receptor cells—the inner and outer hair cells—are located above the basilar membrane. The vast majority of the signals sent from the ear to the brain are from the inner hair cells. The tectorial membrane is located above the hair cells. When waves travel through the cochlea, differences in motion between the tectorial membrane and the basilar membrane cause hairlike stereocilia attached to the top of the hair cells to move. When this happens, filaments that exist between the stereocilia pull open ion channels in the hair cell, allowing potassium ions to enter. This eventually causes the release of a neurotransmitter, most probably glutamate, into the synaptic cleft between the hair cell and an auditory nerve cell. In short, pressure on the hair cells results in the generation and transmission of neural impulses that ultimately travel to the brain.

When a sound wave is transmitted to the cochlea, the mechanical properties of the basilar membrane produce a wave in that membrane that travels down the length of the cochlea, but it has a maximum height in a place that depends on the sound's frequency. High frequencies produce a maximum height near the base (where the oval and round windows are located); low frequencies produce a maximum height at the apex (near the helicotrema). The cochlea's ability to separate similar frequencies is better than would be expected from a passive mechanism. Apparently the outer hair cells, which contain proteins similar to those in muscle cells, somehow act to fine-tune this process by acting as both an amplifier and a filter.

The processing of sound by the brain is not as well understood as that of vision. There are multiple areas in the brain where processing occurs and many interactions between lower and higher centers, as well as between centers on either side of the brain. Like the visual cortex, the auditory cortex contains numerous mappings and numerous specialized cells. Maps in the visual cortex are retinotopic (adjacent areas of the retina are mapped in adjacent areas of the cortex), whereas maps in the auditory cortex (and in other auditory brain areas) are tonotopic (adjacent frequencies are mapped in adjacent areas of the cortex). Cortical cells have been found that respond best to sounds coming from similar positions in space, sound increasing or decreasing in frequency, sounds changing in amplitude, complex sounds such as noise or clicks, or several other qualities (Moore, 2004). Damage to the human auditory cortex does not eliminate the perception of sound, but it can result in loss of higher-order perceptions such as the ability to locate sound sources (Zatorre & Penhune, 2001), or to perceive qualities associated with music (Samson, 1999), or to perceive speech (Boatman, 2006).

LOUDNESS, PITCH, AND LOCALIZATION

In the next section we will consider two basic subjective qualities of sound, loudness and pitch. These qualities are strongly related to the physical quantities of intensity and frequency. Relations exist within these pairs as well as between them. Intensity cannot exist independently of frequency—any sound has both—and the shortest transient change in pressure can be analyzed into multiple frequencies. Within the psychological qualities, loudness depends upon pitch—a very low-pitched tone sounds much less loud than a higher-pitched tone when both have the same energy. Also, changing the loudness of a tone can change its perceived pitch.

Loudness

Loudness is the perceptual quality that is most strongly correlated to the intensity of the sound. Loudness, however, is also related to the frequency of the sound and, to a much lesser extent, the duration, bandwidth, and context of the sound.

Psychological Measures of Loudness

Loudness can only be measured indirectly through the responses of participants in psychophysical studies. There are two widely used units of loudness, the *phon* and the *sone*. The phon is based on the loudness of a 1,000 Hz tone at various sound pressure levels, the sound pressure level equaling the phon value. For example, a 40 dB 1,000 Hz tone has the loudness value of 40 phons. A tone of any frequency that is judged to have the same loudness also has a value of 40 phons. Curves showing equal loudness across frequencies, *equal loudness contours*, show the way that loudness varies across tone frequencies. Figure 25.9 shows the relation between loudness and frequency at 40 and 80 dB. Note that as the tone frequency increases or decreases from 4,000 Hz, the perceived loudness decreases. Note also that at higher sound pressure levels the curve is flatter. If you wished a 30 Hz bass tone to be as loud as a higher frequency tone, you would not have to increase it as much at higher dB levels.

Although the phon is an adequate measure of loudness for many purposes, it is not a ratio scale—80 phons is not twice as loud as 40 phons. In order to have a more psychologically relevant unit, Stanley Smith Stevens (1936) proposed that a new unit be used, the sone. One sone is defined as 40 phons, the loudness of a 1,000 Hz tone at 40 dB (a 40 dB sound is approximately as loud as the normal background noise present in a quiet office). When another

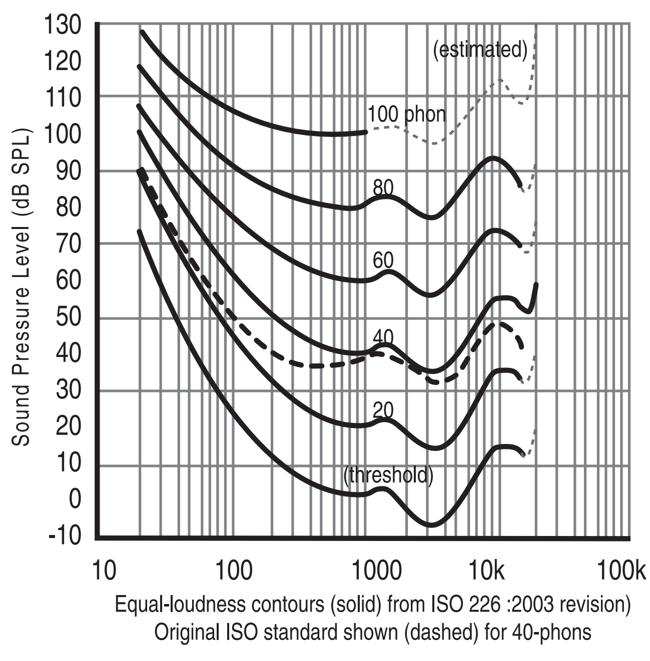

Figure 25.9 Equal loudness contours.

tone's intensity is adjusted so that it is perceived to be twice as loud, that sound is given the value of two sones. If a tone is modified to be half as loud, that tone is given the value of one half sone. This method of scaling loudnesses is called *ratio production* and was often used by Stevens.

Time and Loudness

The auditory system requires time to process loudness. It seems as if the loudness of a sound takes time to get up to its true level. Up to about 200 ms, increasing the length of a sound of a certain intensity produces increased loudness. Below 80 ms, it seems that there is a direct tradeoff between duration and intensity—any combination of the two that leads to the same overall amount of energy is heard as equally loud. The relation beyond 80 ms is more complicated (Moore, 2004)

Duration can also decrease loudness. In the same way that we adapt to the pressure sensations of the clothes we wear, the sensation of loudness undergoes adaptation. There is conclusive evidence that the rate of firing measured in the auditory nerve decreases with time. The psychophysical evidence has been mixed as to the exact details of loudness adaptation, and this decrease in loudness is difficult to notice in everyday life. However, the results of several studies show that a fairly large amount of loudness adaptation occurs (Jones, Weiler, Warm, Dember, & Sandman, 2003; Hellman, Miskiewicz, & Scharf, 1997). A possible real-life example of loudness adaptation is the extreme difference in apparent loudness of a car radio when the volume setting of the night before seems too loud when starting the car the next morning.

Other Loudness Phenomena

Contrast effects, like those in vision, can also occur. The judgment of the loudness of a quiet tone is decreased if that tone is preceded by a louder tone (Garner, 1954). Whether this is more a matter of judgment than sensation is uncertain (Melamed & Thurlow, 1971). An opposite effect can occur when tones of different frequencies and intensities are heard simultaneously. Marks (1994) found that in this situation, given sufficient differences in frequencies, the less intense sound increases in loudness. This is a process of assimilation rather than contrast. Marks suggested that this phenomenon might be due to greater adaptation to the louder tone than the softer tone.

The frequency makeup of noise also influences loudness. If the upper and lower limits of bandpass noise (noise limited to frequencies within certain limits) are extended while the overall energy of the noise remains constant, loudness increases. But this will only occur when the bandwidth is larger than a certain size. Apparently, if the bandwidth is too small, only receptors specific to a certain frequency are affected. When the bandwidth is larger, receptors that encode several frequencies are stimulated and their combined activity produces a greater sensation of loudness. The bandwidth within which the loudness does not change is called the *critical bandwidth* and is an important quantity in many areas of hearing research.

When two separate tones are present, the loudness of their combination depends on how closely their frequencies lie. If the two equally intense tones are very close, the loudness of the tones will alternately increase and decrease as the pressures changes in the tones either add to or subtract from each other. This phenomenon is known as *beating* and can be heard when two guitar strings are tuned to nearly the same note. When the tones are farther apart and the beats disappear, the combined tones gradually increase in loudness until the two tones seem about twice as loud as one of them alone (Zwicker & Fastl, 1990).

Neural Coding of Loudness

Perhaps the most natural way for neurons to carry information about the intensity of a sound is through their firing rate. This explanation seems unlikely, though, when the limited frequency range of neuron action potential rates is compared with the 1 to 10,000,000 range of pressure differences that we can sense.

One possible way of encoding loudness is through an increase in the area of the cochlea stimulated, with a corresponding increase in the number of different neurons responding, as the intensity of the tone increases. However, when the area available for stimulation is limited by the presence of masking noise above and below the frequency of the tone that is being judged, loudness perception is still accurate (Plack, 2005).

The mechanical action of the cochlea along with the amplification of only low-intensity sounds by the action of the outer hair cells act to compress the frequency range needed to be carried by the auditory nerve (in the same way the decibel scale compresses the huge differences in pressure). There is also evidence that the cochlear nucleus, the first brain area to process information coming from the cochlea, acts to compress the neural loudness signal (Zeng & Shannon, 1994).

Pitch

Pitch is the perceptual quality that is most strongly correlated with the frequency of the sound. Psychologists and others studying audition have defined several types of pitch. There is *tonal pitch,* the type of pitch that we normally think of as pitch, related to the sinusoidal components of the sound. *Periodicity pitch* requires only that a sound (or sound gap) occur in a regular manner. *Tone height* is the simple judgment that one sound is higher or lower in pitch than another. *Tone chroma,* or musical pitch, includes judgments of octaves and other musical intervals. Melodies require tone chroma. Until the maturation of the psychological study of music, psychologists generally investigated tone height. Currently, some psychologists such as Plack (2005) believe that the term *pitch* should be limited to those sensations that can be used to produce musical melodies.

The fact that pitch does not require a continuous tone has been known for many years. One of the earliest apparatuses for producing high-frequency sound was Savart's wheel, used in the first part of the 19th century. A thin card was held against the rotating teeth of a rapidly spinning gear. Although the sound produced was very different from that of a sine-wave tone or of a tone produced by a musical instrument, a definite pitch, usable to produce melodies at low enough frequencies, could be heard that matched the frequency of the gear teeth striking the card (Boring, 1942). If, in your youth, you ever placed a card or balloon so it hit the spokes of a bicycle wheel, you have experienced something similar to Savart's wheel. In a more recent example, Miller and Taylor (1948) used regular interruptions of white noise to produce stimuli that had the quality of pitch. Plack believes that all pitch is periodicity pitch. The sensation of pitch that is produced by a sinusoidal tone is merely a special example of periodicity pitch, where the period is that of some portion of the waveform.

Pitch Sensitivity

People are very sensitive to changes in pitch. Below 1,000 Hz, people can reliably distinguish pure tones less than 1 Hz apart (some sources double the needed difference). As frequency increases, the just noticeable difference increases, but even at 10,000 Hz differences of 100 Hz can be distinguished. The Weber's fraction for pitch discrimination is approximately 1/333. This is considerably better than that for loudness discrimination (1/11) or visual brightness discrimination (1/62). The ability to hear higher frequencies decreases markedly with age, a condition called *presbycusis*.

Intensity and Pitch

As mentioned previously, the intensity of a sound can alter its pitch, but the amount of change seems to vary a great deal among observers. Stevens found rather large differences in a participant whom he labeled as a good responder. Increasing the intensity led to an increase in pitch of about 10% for high frequencies, a decrease in pitch of about the same extent for low frequencies, and no change for tones at 3,000 Hz. Later studies found much fewer differences, and those were only for certain participants (Gelfand, 1998).

Measuring Pitch

Because of the strong relation between pitch and frequency, it is not uncommon for studies to use frequency (Hz, cycles per second) to refer to pitch. Doing so may be useful in many cases, but our perception of pitch can vary considerably from the stimulus frequency. One of the first attempts to develop a psychological scale for pitch was done by Stevens and others in 1937 using methods similar to those they used in developing the sone scale of loudness, described in a previous section. A sinusoidal tone of 1,000 Hz at 40 dB was assigned the value of 1,000 mels. Participants then used the method of fractionalization—for example, changing the frequency until the tone appeared to have twice the original pitch, giving that pitch the value of 2,000 mels. The resulting mel scale showed a subjective compression of the frequency values. For example, the 2,000 and 4,000 Hz stimuli were given the mel values of approximately 2,000 and 3,000 mels (Siegel, 1965). Stevens's data deviated considerably from a power curve when the stimuli were much higher or much lower than 1,000 Hz. A replication by Siegel suggested that the reason for this could have been the age of the participants (the older the participant, the less well he or she can hear high-frequency tones) and the sound insulation in the listening room, which absorbed high frequencies more than low frequencies. This finding, as well as one by Greenwood (1997) that Stevens's data was potentially biased by the participants' listening to ascending series of stimuli more than descending series, makes it clear that extreme care must be taken when doing psychophysical experiments.

A more recent pitch scale is the *bark* scale, based on critical bands. As you may recall from the section on loudness, extending the range of noise (without changing its energy) within a critical bandwidth doesn't alter apparent loudness, but expanding it beyond this bandwidth increases the loudness. Also, two tones that are separated by less than the critical bandwidth result in the perception of *beats*, an increase and decrease of their loudness. Once the tones are separated by more than the critical band, the two tones can be heard separately without any perception of roughness. Critical bands exist around all frequencies, therefore they exist as overlapping bands. However, if frequencies are chosen so that their bands do not overlap, humans have approximately 24 critical bands. These bands increase in width as frequency increases and are proportional to our ability to discriminate pitch. The scale of critical bands ranging from 1 to 24 is the bark scale. Its shape is reasonably close to that of the mel scale (Buser & Imbert, 1992). The erb scale (standing for the *equivalent rectangular bandwidth*) is similar to the bark scale, but it measures the intervals in a way that is less subject to error.

Musical Pitch

Musical pitch, or tone chroma, is qualitatively different than tone height, the sort of pitch that is measured by the mel and bark scales. Musical tones are measured in scales of repeating notes, where musically similar tones are an octave apart. Middle C on the piano sounds qualitatively similar to all the other Cs on the keyboard. Musical intervals such as octaves and fifths are directly related to frequencies. Raising a note by an octave doubles the frequency. Raising a note by a fifth, going from an A to an E, increases the frequency by 3/2. Raising it by a fourth increases the frequency by 4/3. The actual tuning of musical instruments does not match these perfect ratios in order that transpositions from one key to another on an instrument with fixed notes, such as an organ or a piano, will retain these intervals as closely as possible. Preference for intervals such as an octave or a fifth seems to occur in cultures having very different musical traditions and are therefore probably innate.

Musical instruments produce complex tones, not simple sine-wave tones. Even a simple plucked string produces not just a single frequency, the *fundamental*, but also multiples of that frequency, called *harmonics*. The quality of a note that can distinguish one instrument from another is called its *timbre* (often pronounced "tamber"). The relative strength of the harmonics is a major component of timbre. When two notes are played together, the overtones of those notes can produce either *consonance* or *dissonance*. Dissonance occurs when the overtones of the notes are closer than a critical bandwidth apart and, as we mentioned before, they combine to produce an audible roughness.

Because our auditory system is sensitive to the periodicity of sound, the fundamental can be heard even if only the overtones are present. For example, when the lowest key on a piano is struck, the fundamental is clearly heard even though it is extremely weak in comparison to the overtones. As another example, when music is played through a very small speaker, bass notes are heard even if the speaker is physically incapable of producing them. The reason why this occurs can be seen when the combination of several overtones is graphed in the time domain.

It is not necessary that a tone be heard in order that pitch may be heard. If a 500 Hz sine wave is produced that is only 5 thousandths of a second long, it is heard as a click rather than a tone. However, if it is compared to a 1,000 Hz tone of the same length, the difference in pitch is clearly perceptible. The sound must be at least a few complete wavelengths long for this pitch to be heard.

Neural Coding of Pitch

Pitch seems to be encoded in the auditory nerve by at least two related mechanisms. The first code, the *rate-place code*, is based on the fact that the maximum displacement of the traveling wave in the cochlea depends on the frequency of the tone. If only a single frequency is present, the maximum activity of the hair cells occurs at the point of the cochlea where this wave is at a maximum. The greater the intensity of the tone, the more rapidly the neurons associated with those hair cells fire. Frequency is encoded by the particular set of neurons that are firing most rapidly. There is disagreement over whether this mechanism is sufficient to explain frequency sensitivity and qualitatively different types of pitch. Many auditory scientists believe that an additional mechanism is needed. Axons in the auditory nerve lock into a particular phase of the sound wave, a single portion of the waveform. Thus, the firing rate of the neuron can encode not only the intensity of a sound, but also its frequency. Limitation in firing rates of neurons prevent a direct encoding of moderately high frequencies, but the combined output of several neurons, even if some waves are missed, contains sufficient information to encode frequencies up to about 5,000 Hz in a very accurate manner. This timing code may be related to musical pitch in that frequencies over 5,000 Hz cannot be classified as particular musical notes.

The *missing fundamental* that can be heard when overtones are present is not generally produced in the cochlea, although it can be if the harmonics are sufficiently intense (Plack, 2005). Houtsma and Goldstein (1971) found that melodies with notes including only two randomly chosen harmonics were easily recognizable when the harmonics were presented to the same ear. They are also heard when the two harmonics are presented *dichotically*, one harmonic to each ear. This suggests that the pitch of the missing fundamental is computed in the brain rather than in the cochlea. Further evidence of the central processing of the missing fundamental can be seen when the harmonics are presented successively rather than simultaneously and the missing fundamental is still perceived.

Localization

The sense of hearing allows us to find, track, and identify objects that emit sounds. The ability to point to the object is primarily due to the fact that we have two ears and can compare the intensity and timing of the sounds they detect. Secondary cues include the pattern of intensities of different frequencies that occur depending on the sound's direction because of the shape of our external ear, head, and shoulders. Locating a set of sounds in a particular position in space helps us judge them to be coming from a single source. Finally, the pattern of sound that we receive allows us to judge the sort of space that we are in.

In order to describe how we are able to use our ears to locate objects in space, it is necessary to have a standard description of the space around us. The angle of an object to our left or right is known as the *azimuth* (from *as-sumūt*, "the ways" in Arabic). This angle is measured in a clockwise direction, with a 0-degree azimuth directly in front of us, a 90-degree azimuth directly to our right, a 180-degree azimuth directly behind us, and a 270-degree azimuth directly to our left. The up-and-down direction is measured in degrees of elevation. A 0-degree elevation is at the horizon, a 90-degree elevation is straight up, and a –90-degree elevation is straight down. Using these two angles, given a particular starting location and orientation, we can describe any direction. If we add a measure of distance, we can describe any location.

Our ability to use audition to locate the direction of objects varies with azimuth and elevation, as well as with the type of sound. The minimal audible angle, the smallest difference in direction that we can reliably detect, is at 0 degrees azimuth and elevation and at frequencies below 1,000 Hz. The ability to locate sounds is also fairly good above 4,000 Hz, but it is inferior between those two frequencies.

The obvious cause of our ability to locate sounds is the fact that we have two ears separated by about 5.5 inches of space and pointing in opposite directions. If a sound comes from the right, our right ear hears it first and most intensely. If the sound approaches from directly in front, there is no difference in intensity or timing. However, if these were the only two factors used, we would be unable to distinguish sounds directly ahead from those directly behind. Our ability to do so is much better than chance, so other sources of information must exist.

The sense of hearing is our most sensitive gauge of time. We are able to judge that a sound source has changed position using timing cues when the time difference between the ears is about 15 microseconds (millionths of a second). When a continuous sound, such as a sine-wave tone, is heard, something other than the time that the sound first affects the ear must be used if location is to be judged. Recall from the section on pitch that the neurons directly affected by the hair cells tend to fire at one particular point in the sound wave. Because of this, differences due to the *phase* of the sound wave can be detected by each ear and compared. If the right ear detects the wave's rising pressure and the left ear detects the peak pressure, the brain has information suggesting that the sound came from the right. Interaural time differences are most effective at lower frequencies. If the frequency is too high, the wavelength is short enough that the phase information becomes ambiguous. By the time the sound travels from one ear to the other, several cycles of the wave may have occurred.

Interaural intensity differences only occur at higher frequencies. If the frequency is too low and the wavelength too long, the head does not effectively block the sound. However, high-frequency sounds are blocked quite well. A 6,000 Hz tone coming from one side of the head decreases by about 20 dB (Moore, 2004). The previously mentioned decrease in accuracy between 1,000 and 4,000 Hz is probably due to switching between time and intensity information. Most sounds encountered in everyday life contain a combination of lower and higher frequencies and both mechanisms are active.

Discriminating elevation and front-to-back location is accomplished by the filtering caused by the shape of the outer ear and head. When the shape of the outer ear is modified by filling in its hollows, the ability to distinguish these directions is greatly reduced. Binaural recordings, recorded with microphones placed in the ears of a replica of a person's head, provide markedly improved realism. The location of the instruments in an orchestra can be more accurately perceived when one listens through headphones. Unfortunately, the realism disappears if the listener's head is turned.

The perception of distance is much inferior to that of direction. Loudness can provide distance information if the sound is a familiar one, such as a person's voice. For a complex sound, the higher frequencies tend to diminish with distance somewhat more than low frequencies because of absorption of high frequencies by the atmosphere.

The judgment that a complex sound comes from a single source is essential for the organization of the auditory world. Sounds that appear to come from the same spatial position are often perceived to be coming from the same source. When position cues are diminished through the use of hearing aids, it becomes more difficult to separate someone's voice from other noises in the environment or from other voices. Identifying a single instrument when listening to an orchestra can be done even if no positional information is available. The pattern of harmonics and the way in which the intensity of the sound changes over time, factors that produce the timbre of the instrument, are sufficient for its identification.

In any enclosed space (except for an anechoic chamber), a good part of the sound that reaches us from a source comes not directly from that source but from reflections from the walls. This indirect sound reaches us from varying directions and with varying delays, yet only one sound is heard coming from the direction of the object producing it. The fact that we do not separately hear the echoes is due to what is called the *precedence effect*. The first sound that reaches us suppresses later echoes. Depending on the type of sound, a gap of up to 50 microseconds between the initial sound and the reflected sound produces the percept of a single sound (Gelfand, 1998). Even though only one sound is heard, the echoes provide information concerning the size of the room; the greater the time difference, the larger the room is perceived.

REFERENCES AND FURTHER READINGS

Aristotle. (350 BCE). *On the Soul*. Retrieved April, 7, 2007, from classics.mit.edu/Aristotle/soul.html

Boatman, D. F. (2006). Cortical auditory systems: Speech and other complex sounds. *Epilepsy and Behavior*, *8*, 494–503.

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century.

Burtt, H. E. (1917). Auditory illusions of movement—a preliminary study. *Journal of Experimental Psychology*, *2*, 63–75.

Buser, P., & Imbert, M. (1992). *Audition*. Cambridge, MA: Bradford.

Cusack, R, & Carlyon, R. P. (2003). Perceptual asymmetries. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 713–725.

Garner, W. R. (1954). Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, *48*, 218–224.

Gelfand, S. (1998). *Hearing*: *An introduction to psychological and physiological acoustics* (3rd ed.). New York: Marcel Dekker.

Greenwood, D. D. (1997). The mel scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distance and equal frequency ratios. *Hearing Research*, *103*, 199–224.

Hellman, R., Miskiewicz, A., & Scharf, B. (1997). Loudness adaptation and excitation patterns: Effects of frequency and level. *Journal of the Acoustical Society of America*, *101*, 2176–2185.

Houtsma, A. J. M., & Goldstein, J. L. (1972). The central origin of the pitch of complex tones: Evidence from musical interval recognition. *The Journal of the Acoustical Society of America*, *51*, 520–529.

Jones, A. T. (1937). *Sound, a textbook*. New York: Van Nostrand.

Jones, K. S., Weiler, E. M., Warm, J. S., Dember, W. N., & Sandman, D. E. (2003). Loudness adaptation: Fact or artifact? *The Journal of General Psychology*, *130*, 341–358.

Luce, R. D. (1993). *Sound and hearing*. Hillsdale, NJ: Erlbaum.

Marks, L. E. (1994). "Recalibrating" the auditory system: The perception of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, *20,* 382–396.

Melamed, L. E., & Thurlow, W. (1971). Analysis of contrast effects in loudness judgments. *Journal of Experimental Psychology*, *2*, 268–274.

Miller, G. A., & Taylor, W. (1948). The perception of repeated bursts of noise. *Journal of the Acoustical Society of America*, *20*, 171–182.

Moore, B. (2004). *Psychology of hearing*. London: Elsevier.

Plack, C. J. (2005). *The sense of hearing*. Mahwah, NJ: Erlbaum.

Rich, G. J. (1916). A preliminary study of tonal volume. *Journal of Experimental Psychology*, *1*, 13–22.

Samson, S. (1999). Musical function and temporal lobe structures: A review of brain lesion studies. *Journal of New Music Research*, *28*, 217–228.

Siegel, R. J. (1965). A replication of the mel scale of pitch. *The American Journal of Psychology*, *78*, 615–620.

Stevens, S. S. (1936). A scale for the measurement of a psychological magnetite: Loudness. *Psychological Review*, *43*, 405–416.

Stevens, S. S. (1937). A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, *8*, 185–190.

Zatorre, R. J., & Penhune, V. B. (2001). Spatial localization after excision of human auditory cortex. *The Journal of Neuroscience*. *21*, 6321–6328.

Zeng, F. G., & Shannon, R. V. (1994). Loudness-coding mechanisms inferred from electric stimulation of the human auditory system. *Science*, *264*, 564–566.

Zimbardo, P. G., Andersen, S. M., & Kabat, L. G. (1981). Induced hearing deficit generates experimental paranoia. *Science*, *212*, 1529–1531.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics: Facts and Models.* Berlin: Springer.

26

Somatosensory Systems

Sharon E. Guttman
Middle Tennessee State University

Stephen F. Davis
Texas Wesleyan University

The *somatosensory systems* receive and process somatic, or bodily, information. Like all the other senses (vision, hearing, taste, and smell), the somatosensory systems reveal qualities of stimuli present in the external world (*exteroceptive* information). However, the somatosensory systems play an additional role, unique among their counterparts: They provide *interoceptive* information—data regarding internal conditions, such as core temperature and internal pain.

As this discussion implies, somatosensation involves the operation of several distinct systems. In this chapter, we discuss several of these systems, including touch, pain, proprioception (i.e., information about bodily position), and temperature regulation. As you read further, however, keep in mind that these systems, though distinct, interact with one another. Together, the various somatosensory systems offer a rich description regarding states of both the internal and external environments.

TOUCH

If you are like most individuals, you are acutely aware of the importance of your vision and hearing for everyday life—the thought of blindness or deafness is devastating. But when is the last time you considered the importance of your sense of touch? Upon awakening in a dark room in the middle of the night, your awareness of your surroundings depends primarily on tactile sensations—the feel of the linens against your skin, or the furry dog underfoot as you attempt to rise from the bed. Imagine finding your ringing cell phone in the bottom of your purse or schoolbag in the absence of touch. By the time you searched through the clutter using vision alone, the ringing would have stopped. The sense of touch reveals not only the presence of an object but also information about its shape, size, firmness, and texture, all important qualities for our interactions with objects in the environment.

The sense of touch may serve another important purpose as well. Though we often consider vision to be the most *informative* sense, we probably should consider touch to be the most *reliable* sense. Imagine reaching toward a seen object, only to feel nothing but thin air. Upon which of your senses would you rely? In the face of such conflict, we typically trust our sense of touch; the information revealed by direct contact with an object (or lack thereof) cannot be easily denied.

Receptors for Touch

Your sense of touch involves several types of receptors, called *mechanoreceptors,* located in the skin. The mechanoreceptors are so named because they respond directly to mechanical stimulation: pressure or deformation of the skin.

Two types of mechanoreceptors, the Meissner corpuscles and the Merkel disks, reside immediately beneath the skin's surface. These two receptor types have small, punctate receptive fields: They respond to touch information over a small, distinct area of the skin. As such, these receptor types are responsible for processing fine details of a tactile stimulus, including texture. To appreciate the importance of these receptors, consider that the density of Meissner corpuscles in the fingertips declines steadily from 40–50 per square millimeter of skin during late childhood to around 10 per square millimeter by age 50. This decline predicts accurately the loss of sensitivity to detailed tactile information that older people experience (Thornbury & Mistretta, 1981).

The other two types of mechanoreceptors, the *Pacinian corpuscles* and the *Ruffini endings*, inhabit deeper positions in the skin. These receptors have large, diffuse receptive fields: They respond to touch information over a much larger, indistinct region, and provide more general, "big picture" information about the nature of a touch stimulus. For all four receptor types, receptive field size—the area of skin over which a touch stimulus can produce a response—varies over the body. Receptive field sizes are much smaller, and thus provide more precise information, in regions of the body of evolutionary importance for processing touch information: the fingertips and lips.

It should be noted that the mechanoreceptors also differ in their rate of adaptation. The Merkel's disks and Ruffini endings adapt slowly; they respond relatively steadily to continuously applied pressure. By contrast, both the Meissner's corpuscles and the Pacinian corpuscles adapt very rapidly; they respond vigorously when a tactile stimulus first contacts the skin and when the stimulus is removed, but little in between. If you are wondering about the relevance of fast and slow adaptation, consider the following situation. It's a cold winter morning and you hurry to pull on your favorite sweatshirt. For a very brief time you may be aware of the feel of your sweatshirt against your skin, but you soon become unaware of any pressure of your clothing (unless you consciously focus on it). This scenario reflects the response of the rapidly adapting receptors. Initially, both slowly and rapidly adapting receptors respond; then, after a brief period (e.g., 300–600 milliseconds), the Meissner and Pacinian corpuscles adapt and your tactile experience changes. The importance of rapidly adapting touch receptors can best be experienced by closing your eyes and trying to identify an object by only touching it. Can you identify the object without keeping your hands in constant motion? No—you need the constant motion to keep the rapidly adapting touch receptors responding and providing you with information.

Haptic Perception

This discussion highlights a general principle of perception: To glean the best information regarding an object or event, we actively explore it using whichever senses are available to us. Think about what you would do if asked to describe, in detail, the appearance of an object sitting in front of you. Most likely, you wouldn't simply glance at the object; you would systematically move your eyes over its surface, examining every detail. Similarly, to fully appreciate an object's tactile properties, you move it around with your hands, exploring it in a variety of ways. Researchers refer to active exploration using the sense of touch as *haptic perception*.

Lederman and Klatzky (1987, 1990) discovered that people engage in a series of predictable, ritualized actions, deemed *exploratory procedures*, when examining objects with their hands. Each exploratory procedure reveals different types of information about the object in question. For example, "lateral motion" of the fingers across an object's surface yields texture information, whereas "contour following" provides data regarding the object's shape. As long as use of these and other exploratory procedures is not constrained, people can identify objects by touch alone with remarkable speed and accuracy (Klatzky, Lederman, & Metzger, 1985; Lederman & Klatzky, 2004).

As just suggested, haptic perception yields much more information about the nature of an object than passive touch. Attempts to identify an object by simply grasping it, with no exploration, often lead to gross errors in perceptual judgment (Rock & Victor, 1964). No doubt, the superiority of haptic perception must be attributed, in part, to the ongoing activity of the detail-oriented, rapidly adapting touch receptors. However, haptic perception trumps passive touch even after taking adaptation into account; moving the object across your stationary hand, which fully engages rapidly adapting receptors, does not produce the same level of perceptual knowledge as active exploration. The reason for the difference in utility between active and passive touch can be simply explained when you recall that all of the somatosensory systems interact with one another. Tactile input alone provides ambiguous data regarding the nature of an object. Feeling a sharp point, for example, reveals little about an object if you don't know where the point is positioned relative to other features, or what actions your fingers made to discover the point. To make sense of what your mechanoreceptors signal, you must take into account the relative positions and movements of your fingers. In other words, haptic perception involves integrating tactile information with sensory information regarding *proprioception* and *kinesthesis*. We return to these aspects of somatosensation later in the chapter.

Dermatomes

Touch receptors synapse with neurons that carry the touch information to the spinal cord and then to the brain for processing. It is noteworthy that the neurons from distinct, identifiable "strips" of the body enter the spinal cord together and ascend to the brain as a group. Researchers call the bodily strips that send groups of neurons coursing upward to the brain *dermatomes*. The dermatomes on the left side of the body mirror the dermatomes on the right side of the body. Damage to the axons entering the spinal cord at a single level can cause a loss of tactile sensation from a single dermatome of the body.

Ascending Somatosensory Pathways

Two major neural pathways ascend the spinal cord and carry somatosensory information to the brain. The *dorsal-column medial lemniscus system* carries touch information and ascends in the dorsal (back) portion of the spinal cord. The *anterolateral system*, also called the *spinothalamic tract*, conveys information about pain and temperature (discussed later) and ascends in the anterior (front) portion of the spinal cord. It is noteworthy that the dorsal-column medial lemniscus system is a *contralateral system*; that is, the neurons from the left side of the body cross from the left side of the spinal cord to the right side and send their information to the right hemisphere of the brain. On the other hand, the anterolateral system is an *ipsilateral system*. That is, the majority of the neurons in this system do not cross as they ascend the spinal cord; for example, information originating on the left side of the body is processed in the left hemisphere of the brain.

Cortical Processing of Somatosensory Information

What happens when the somatosensory information carried by the two ascending pathways reaches the brain? To answer this question, we need to examine the classic research conducted by neurosurgeon Wilder Penfield and his colleagues. During brain surgery, these investigators electrically stimulated the brains of their human patients, who were awake and could verbalize, and recorded the responses that these stimuli produced (see Penfield & Boldrey, 1937; Penfield & Rasmussen, 1950). Upon stimulation of the frontmost part of the parietal lobe, the *postcentral gyrus*, patients reported experiencing tactile sensations. As the region of the brain most central to somatosensory processing, the postcentral gyrus is commonly referred to as *primary somatosensory cortex*.

Further investigations by Penfield and his colleagues revealed two important characteristics of primary somatosensory cortex. First, different regions along the postcentral gyrus correspond to different parts of the body. Adjacent regions of the body (e.g., the hand and the arm) receive representations in adjacent regions of cortex. Second, the size of areas on this cortical map, which is consistent from person to person, varies according to the relative sensitivity of specific areas of the body to tactile stimulation. The greater the density of mechanoreceptors in a given region, the more sensitive that region, and the larger the cortical representation tends to be. Areas like the fingers and lips, which are especially important for adaptation to and interaction with the environment, occupy greater cortical space than do areas of less adaptive importance, such as the feet and back.

What sort of experience would you expect if the somatosensory cortex fails to process tactile information completely or adequately? Although rare, failure of the brain to process touch stimuli can result in *astereognosia*, also called *tactile agnosia*. Individuals suffering from astereognosia experience basic tactile sensations, but cannot identify objects using the sense of touch.

Plasticity in Somatosensory Cortex

As we've seen, the cortex devotes considerable processing capacity to areas of the body of evolutionary importance for tactile exploration: in humans, the hands and the mouth. This finding implies that, over time, the brain has adapted to the needs of the organism. Importantly, however, adaptation also occurs on a much shorter timescale. Converging evidence suggests that the brain shows remarkable plasticity—flexibility in its representation—with regard to somatosensory processing.

Work by Merzenich and colleagues first demonstrated that increased input to a particular region of skin could cause expansion of the cortical representation of that body region. In monkeys, three months of increased tactile stimulation of a single finger resulted in a vast expansion of the already large cortical representation of the digit in question (Buonomano & Merzenich, 1998). In humans, it has since been demonstrated that the somatosensory cortices of musicians who play stringed instruments contain unusually large representations of the overstimulated fingers of the left hand (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995).

Blind individuals who read Braille with their hands provide further insight into the plasticity of somatosensory representations. Consistent with the research described above, the brains of blind individuals contain significantly larger representations of the dominant Braille-reading finger(s) than of any other fingers (see Hamilton & Pascual-Leone, 1998, for a review). Furthermore, functional imaging studies demonstrate that, in congenitally blind individuals, Braille reading activates areas of the occipital lobe normally used for visual processing (Sadato et al., 1996). When transcranial magnetic stimulation (TMS) momentarily disrupts processing in "visual" processing areas, blind individuals remain aware of the tactile stimulation induced by touching the Braille letters, but have difficulty identifying the letters being touched; no such effects occur in sighted people (Cohen et al., 1997; Kupers et al., 2007). Thus, it appears that somatosensory processes can "hijack" otherwise unused areas of the brain.

What happens when the unused brain tissue actually resides in the somatosensory cortex itself? Under extreme circumstances, such as an automobile accident, a natural disaster, or warfare, doctors may need to amputate a limb.

A vast majority of amputees experience a *phantom limb*; their missing arm or leg continues to “feel” as if it’s still present (Melzack, 1992). The phenomenon, documented for centuries, can best be explained as the perceptual consequence of the lingering brain tissue that represents the amputated limb; when this tissue becomes active, amputees experience sensation in the phantom. More recently, Ramachandran and colleagues (e.g., Ramachandran & Hirstein, 1998) discovered a number of patients who, when their faces are touched, experience not only sensation in their faces but also referred sensation in the phantom arm. Though puzzling, this finding can be explained when we note that the representation of the face within somatosensory cortex lies immediately adjacent to the representation of the hand. Over time, the cortical map adapts, such that tactile input to the face activates not only the face region but also the idle tissue nearby.

Approximately half of all amputees also experience pain from the phantom limb. The exact cause of this pain remains unclear. Some researchers have speculated that this pain is caused by some type of irritation of the nerves at the point where the limb was amputated. However, if this explanation is correct, then damaging the neural pathway that carries information from the point of amputation to the cortex, thus leaving the brain without pain signals to process, should yield a permanent, long-lasting decrease in phantom limb pain. Unfortunately, such surgery has produced, at best, only a very transitory period of relief (Melzack, 1992). We turn now to the general topic of pain.

PAIN

Sometimes the stimuli applied to the skin produce tissue damage. In cases of potential or actual skin damage, in addition to the sense of touch or pressure, we experience pain. Though aversive, the experience of pain serves an adaptive purpose. Acute pain functions as an early warning system; it allows us to act quickly when a stimulus poses an imminent threat to our well-being. More prolonged pain reduces the temptation to touch a damaged region of skin or move an injured joint, supporting the healing process. Some individuals cannot feel pain, despite experiencing basic tactile stimuli, a condition referred to as *idiopathic analgesia*. Though you may be tempted to believe that such individuals are lucky, they actually have a reduced life expectancy.

In their study of the sensation we call pain, researchers have unearthed several intriguing research questions. For example, you have heard athletes talk about “playing through” the pain of a major injury that they have suffered. How do humans use emotional or cognitive processes to completely suppress or block pain? Researchers (e.g., Roland, 1992) also have been puzzled by the fact that when the brain processes information received from the pain receptors, this activity is not localized to discrete areas in the same manner as the processing of touch information in the primary somatosensory cortex. We discuss research findings related to these, and other, questions in this section.

Pain Receptors, Nerves, and Neurotransmitters

The somatosensory receptors responsible for pain are known as *nociceptors*. Different nociceptors react to different types of painful stimuli: sharp touch, extremes of temperature, and chemical stimulation. The body’s primary nociceptors can be found among free nerve endings located just beneath the skin’s surface and in the underlying layer of fat.

Two specialized types of nerve fibers carry pain messages from the nociceptors to the spinal cord. Messages regarding strong touch (e.g., a kick to the shin) or intense heat travel along fast-conducting, Aδ nerve fibers. Activation of these fibers results in a sense of sudden, acute pain, such as when you accidentally touch a hot stove. This kind of pain serves a protective function, causing you to withdraw from the painful stimulus. Messages regarding less intense heat or chemical stimulation—including the neurochemicals released by our own bodies upon injury to the skin— travel along more slowly conducting, C nerve fibers. If your C fibers become active, you might experience a more prolonged, “achy” pain, such as that associated with an injured back. These two types of nerve fibers work together in forming the pain experience. Imagine hitting your thumb with a hammer; you will likely experience a sudden, sharp pain upon first impact, followed by a longer-lasting aching. Moreover, the C fibers (and the Aδ fibers, to a lesser extent) become potentiated upon injury, such that normally innocuous stimuli (e.g., a gentle touch) feel painful.

Upon reaching the spinal cord, the Aδ and C nerve fibers, when active, release two neurotransmitters: glutamate and substance P. If the pain stimulus is mild and short-lived (i.e., causing Aδ activity only), then only glutamate is released; however, if the pain stimulus is strong (i.e., causing both Aδ and C fiber activity), then *both* glutamate and substance P are released (Cao et al., 1998). DeFelipe and colleagues (1998) experimentally verified that the release of substance P precedes the experience of intense pain. Their study utilized “knockout” mice that lacked receptors for substance P. Because these animals could not react to substance P when it was released, the researchers predicted that the mice would react to intense pain stimuli as if they were mild annoyances. Their results supported this prediction.

Cortical Involvement in Pain

The release of glutamate and substance P in the spinal cord sends impulses coursing up the anterolateral system to the brain. Unlike nonpainful tactile stimuli, which result in localized activity in the primary somatosensory cortex, PET scans (see Chapter 16) indicate that painful stimuli cause distributed activity over many regions of the brain (Talbot et al., 1991). For example, regions of somatosensory cortex (specifically, cortical maps that parallel the representations for basic touch sensation) appear to process the sensory aspects of the pain experience and allow the painful stimulus to be localized. By contrast, pain-related activity in an area of the cortex known as the *anterior cingulate gyrus* plays a role in determining the emotional reaction to pain. The contributions of many other pain-processing regions are not, as yet, well understood.

At this juncture, it should be noted that researchers often speak of the *multimodal nature* of pain. The experience of pain cannot be captured adequately by describing the raw sensation, such as "too hot." Pain has a subjective, emotional component as well. Pain both influences and is influenced by one's mental state. Upon taking this complex, multimodal nature into account, it should not be surprising that a variety of cortical areas play a role in processing pain information.

Gate-Control Theory of Pain

In 1965, Melzak and Wall proposed the influential *gate-control theory of pain*. Importantly, this theory reflects pain's multimodal nature. Specifically, the gate-control theory aims to explain the fact that the experience of pain depends not only on the strength of a stimulus but also on a variety of cognitive and emotional factors.

In the gate-control theory, pain sensations begin with the activation of nociceptors and a message sent along the Aδ and/or C nerve fibers, as described above. However, in order for the pain signals to reach the brain, and thus be consciously experienced, they must pass through open "gates" in the spinal cord.

Two different aspects of the gate-control circuitry can cause the gates to close and block the perception of pain. First, the pain signal's transmission is inhibited when the input from the fast nerve fibers carrying basic (nonpainful) tactile sensations exceeds the input from the slower nerve fibers carrying the pain sensations. Do you immediately rub your knee after you've banged it against the underside of your desk? If so, the gate-control theory may explain why; by activating the mechanoreceptors, you can actually prevent (or at least reduce) the pain signal being sent to your brain. Some researchers also hypothesize that acupuncture reduces pain by stimulating these same touch fibers.

Second, gate-control theory proposes that pain can be blocked by *top-down influences*. That is, messages that originate within the brain descend to the spinal cord and inhibit the transmission of pain signals to the brain. As an illustration of top-down control over pain experiences, numerous studies indicate that patients taught relaxation techniques before surgery require subsequently require lower doses of painkillers. Moreover, there appears to be real truth in the saying that "laughter is the best medicine." Laughter and other positive emotions can significantly reduce the experience of pain (e.g., R. Cogan, D. Cogan, Waltz, & McCue, 1987), apparently through top-down influences over spinal cord pain transmission circuits. For further details on the gate-control theory, see Melzak and Wall (1965).

Descending Pain-Control System

Although our understanding of ascending pain pathways remains in its infancy, research conducted since the proposal of the gate-control theory has successfully identified several specific structures involved in the *descending pain-control system*. Activation of neurons in the *periaqueductal gray*—the gray matter surrounding the cerebral aqueduct—initiates the pain-suppression circuit. The importance of this region first became apparent in the late 1960s, when Reynolds (1969) reported that stimulation of the periaqueductal gray reduces the reaction to painful stimuli (i.e., produces analgesia).

By the late 1970s, Basbaum and Fields (1978; see also Fields & Basbaum, 1984) had proposed a complete descending pain-suppression circuit. Activation of the periaqueductal gray results in stimulation of the *raphe nucleus*, a distinct cluster of neurons located in the depths of the medulla in the hindbrain (brain stem). Stimulation of the raphe nucleus, in turn, *inhibits* pain-sensitive neurons in the spinal cord, blocking the experience of the incoming pain stimulus.

To fully understand the pain-suppression circuit, researchers still needed to identify the neurotransmitter that activated the periaqueductal gray. In 1973, Pert and Snyder had discovered that morphine and other pain-suppressing opiate drugs bind to receptors in this area. This discovery ultimately provided the missing link. Because it seems highly unlikely that the human brain would have developed receptors for an exogenous (external) substance, such as morphine, researchers reasoned that the brain produced it own natural opiates. Two subsequently identified brain peptides, *met-enkephalin* and *leu-enkephalin*, though structurally very different from opiates, do indeed stimulate the brain opiate receptors. The discovery of several natural endogenous (internal) morphines, collectively known as *endorphins*, lent further support to the theory.

Both aversive/painful situations, such as running a marathon, and pleasurable situations, such as sexual activity or listening to your favorite music, can trigger the release of endorphins. In sum, cognitive-emotional stimuli and situations prompt the release of endorphins, activating the descending pain-control system and suppressing the experience of pain.

PROPRIOCEPTION, KINESTHESIS, AND THE VESTIBULAR SENSE

Imagine waking up in the middle of the night and realizing that you have a terrible crick in your neck. The sense of pain typically would be followed immediately by the realization that your body is twisted into a very strange posture that you must immediately adjust. This insight arises from *proprioception,* the perception of the position of the limbs and other body parts in space, and the closely related sense of *kinesthesia,* the perception of limb motion.

Though often overlooked, proprioception and kinesthesia play a critical role in everyday life. As discussed earlier in this chapter, haptic perception—active exploration by touch—requires that tactile information be combined with knowledge of how the fingers are positioned and moving across the object. More generally, proprioception and kinesthesia, together with the vestibular sense (discussed later), provide continuous feedback that allows the brain to fine-tune posture and movements.

Two key receptors for proprioception and kinesthesia, referred to as *proprioceptors,* are the *muscle spindles* and the *Golgi tendon organs.* Each muscle spindle consists of a bundle of four to ten specialized muscle fibers that run parallel to the nonspecialized muscle fibers. The muscle spindles encode changes in muscle length; whenever a muscle stretches, the muscle spindles also stretch, resulting in a message being sent to the spinal cord and ultimately to the brain. Muscles that must perform precise movements, such as those in the hands and the jaw, tend to be more densely populated with muscle spindles than muscles that perform coarser movements. The Golgi tendon organs, found at the junction between the tendon and the muscle, provide complementary information about muscle tension.

The fibers of the muscle spindles communicate directly with motor neurons in the spinal cord. The activity of these proprioceptors initiates the stretch reflex; a rapid stretch of the muscle, detected by the muscle spindles, causes the motor neurons to contract the muscle and return it to the original length. The stretch reflex causes your leg to kick out when your doctor taps below your knee. More important, when a heavy object drops into your outstretched hand, the stretch reflex causes your arm to rise back immediately to its initial position.

Of course, muscle position and tension do not depend solely on reflexive spinal cord control; the brain determines the appropriate posture at any given moment. After reaching the spinal cord, information from proprioceptors ascends to the cortex via the dorsal-column medial lemiscus pathway. Ultimately, proprioceptive information reaches a distinct map within the somatosensory cortex, where its processing leads to awareness of body position. The location of the brain region for proprioceptive analysis is noteworthy for two reasons. First, the regions for proprioceptive and for tactile analysis lie nearby each other, supporting the integration necessary for haptic perception. Second, primary motor cortex resides in the precentral gyrus of the frontal lobe, immediately in front of somatosensory cortex; the proximity of motor and somatosensory cortex no doubt facilitates the brain's use of proprioceptive information to program adjustments in movement.

The Vestibular Sense

Unless you find yourself in an unusual situation, such as trying to walk down a very steep slope, how often have you thought about your sense of balance? "Never" most likely would be your answer to that question. Like proprioception and kinesthesia, the sense of balance—the *vestibular sense*—is one that most people take for granted.

The vestibular system consists of two basic components within the inner ear: the *otolith organs* and the *semicircular canals*. The two otolith organs—the utricle and the sacculus—act as gravity detectors, responding to linear accelerations and reporting the head's orientation in space. The three *semicircular canals* detect rotational accelerations of the head; positioned along three different planes at right angles to one another, the semicircular canals can sense rotation in any direction.

As a part of the labyrinthine network that also includes the cochlea (for hearing), the otolith organs and semicircular canals are filled with a thick fluid that moves as the head moves. This fluid movement causes the hair-like *cilia* and the ends of the vestibular receptors (called haircells) located within these chambers to bend; this receptor bending results in neural messages about head position and movement being sent along the vestibular branch of the eighth cranial to the brain. Several subcortical structures, including the *vestibular nuclei* of the brainstem and the *cerebellum* (which aids in the production of smooth, coordinated movements), play a role in the processing of vestibular information. Cortical regions falling adjacent to primary somatosensory cortex also receive projections from the otolith organs and semicircular canals.

This consideration of the vestibular system offers a useful reminder that our knowledge of the world depends on an interaction among all the senses. From your own experience, you may be aware that conflicts between visual and vestibular input cause dizziness and nausea. For example, most people become carsick not when looking out the window, but when reading. In this example, the nausea-inducing conflict results from the fact that your eyes, which are reading the text but not capturing the passing landscape, suggest that you are stationary whereas your vestibular system registers the movement. Those of you who have taken dance lessons may also have found that spinning in circles causes the most dizziness when you *stop* moving. As long as you continue to spin, your visual and vestibular systems provide complementary information about your movement. However, when you halt, the fluid in your semicircular canals, having built up momentum, continues to move for a few moments; thus, your vestibular system implies that you are still spinning, whereas your visual system argues that you're not.

CONTROLLING BODY TEMPERATURE

Discussions of somatosensory systems often overlook introception, the processing of internal conditions. Though introceptive processes tend to occur automatically, with little or no awareness, they serve a critical function for our survival. Thus, we conclude this chapter with a brief discussion of a representative introceptive process, the control of body temperature.

Fish and reptiles are *poikilothermic*—their body temperatures match the surrounding environment. On the other hand, birds and mammals are *homeothermic*—they have physiological mechanisms that allow them to adjust their body temperature in order to maintain optimal activity levels. Multiple mechanisms exist to achieve either additional warmth or cooling, as necessary. For example, if additional warmth is needed, people and animals (a) shiver, (b) huddle together to share body heat, and/or (c) fluff their fur (furry mammals), hair (humans), or feathers (birds) to improve insulation; when all else fails, the system diverts blood flow from the skin to the vital internal organs, thus protecting critical systems. By contrast, if it's too hot and body temperature needs to be reduced, people and animals can (a) sweat, (b) reduce activity, and/or (c) route more blood to the skin in order to cool it.

Research indicates that the *hypothalamus*, a major regulatory center for numerous survival behaviors, plays a crucial role in temperature regulation. The hypothalamus is a small structure, composed of a number of subdivisions or nuclei, located at the base of the brain. Heating or cooling the *preoptic area* of the hypothalamus triggers the appropriate temperature-adjusting bodily reactions, suggesting that this region governs temperature regulation. Cells in the preoptic area receive neural input via the anterolateral pathway from *thermoreceptors* (temperature receptors) in the skin, as well as monitor their own temperature. Animals with surgical damage to the preoptic area cannot regulate their bodily temperatures.

SUMMARY

In this chapter, we discussed how somatosensory systems provide a plethora of information about the world in which we live, all of it crucial to our survival. Tactile input, especially when gained through active exploration, allows us to make reliable and accurate determinations about the nature of the objects within our reach. Pain, though unpleasant, signals that tissue damage has or is about to occur, protecting us from more severe injury. Proprioceptive, kinesthetic, and vestibular information provides continuous feedback about the states of our bodies, facilitating coordinated, finely tuned movements and contributing to haptic perception. Finally, interoceptive processes, such as temperature regulation, allow our bodies to maintain optimal levels of activity. When combined with the perceptual information provided by the other senses, somatosensory processing allows us to achieve a remarkable understanding of the world around us.

REFERENCES AND FURTHER READINGS

Basbaum, A. I., & Fields, H. L. (1978). Endogenous pain control mechanisms: Review and hypothesis. *Annals of Neurology, 4,* 451–462.

Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience, 21,* 149–186.

Cao, Y. Q., Mantyh, P. W., Carlson, E. J., Gillespie, A. M., Epstein, C. J., & Basbaum, A. I. (1998). Primary afferent tachykinins are required to experience moderate to intense pain. *Nature, 392,* 390–394.

Cogan, R., Cogan, D., Waltz, W., & McCue, M. (1987). Effects of laughter and relaxation on discomfort thresholds. *Journal of Behavioral Medicine, 10,* 139–144.

Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Falz, L., Dambrosia, J., et al. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature, 389,* 180–183.

DeFelipe, C., Herrero, J. F., O'Brien, J. A., Palmer, J. A., Doyle, C. A., Smith, A. J. H., et al. (1998). Altered nocioception, analgesia and aggression in mice lacking the receptor for substance P. *Nature, 392,* 394–397.

Fields, H. L., & Basbaum, A. I. (1984). Endogenous pain control mechanisms. In P. D. Wall & R. Melzak (Eds.), *Textbook of pain* (pp. 142–152). Edinburgh: Churchill Livingstone.

Hamilton, R. H., & Pascual-Leone, A. (1998). Cortical plasticity associated with Braille learning. *Trends in Cognitive Sciences, 2,* 168–174.

Klatzky, R. L., Lederman, S. J., & Metzger, V. (1985). Identifying objects by touch: An "expert" system. *Perception & Psychophysics, 37,* 299–302.

Kupers, R., Pappens, M., Maertens de Noordhout, A., Schoenen, J., Ptito, M., & Fumal, A. (2007). rTMS of the occipital cortex abolishes Braille reading and repetition priming in blind subjects. *Neurology, 68,* 691–693.

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology, 19,* 342–348.

Lederman, S. J., & Klatzky, R. L. (1990). Haptic classification of common objects: Knowledge-driven exploration. *Cognitive Psychology, 22,* 421–459.

Lederman, S. J., & Klatzky, R. L. (2004). Haptic identification of common objects: Effects of constraining the manual exploration process. *Perception & Psychophysics, 66,* 618–628.

Melzack, R. (1992, April). Phantom limbs. *Scientific American, 266,* 120–126.

Melzak, R., & Wall, P. D. (1965). Pain mechanisms: A new theory. *Science, 150,* 971–979.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representations in cerebral cortex of man as studied by electrical stimulation. *Brain, 60,* 389–443.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man: A clinical study of the localization of function.* New York: Macmillan.

Pert, C. B., & Snyder, S. H. (1973). The opiate receptor: Demonstration in nervous tissue. *Science, 179,* 1011–1014.

Reynolds, D. V. (1969). Surgery in the rat during electrical analgesia induced by focal brain stimulation. *Science, 164*, 444–445.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science, 143,* 592–596.

Roland, P. (1992). Cortical representation of pain. *Trends in Neurosciences, 17*, 3–5.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G., et al. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature, 380,* 479–480.

Talbot, J. D., Marrett, S., Evans, A. C., Meyer, E., Bushnell, M. C., & Duncan, G. H. (1991). Multiple representations of pain in human cerebral cortex. *Science, 251,* 1355–1358.

Thornbury, J. M., & Mistretta, C. M. (1981). Tactile sensitivity as a function of age. *Journal of Gerontology, 36,* 34–39.

27

Perception

Lauren F. V. Scharff
Stephen F. Austin State University

Imagine that you and a friend are taking a walk through a wooded area. As you walk, you become aware of the many stimuli in your surrounding environment: the sound of old leaves shuffling under your feet and birds singing high in the trees, the sight of branches that hang in front of your path, the different touch textures of tree bark when you lean on the trees for support as you go down a steep incline, the sharp pain in your finger as you move a briar vine and accidentally grasp a thorn, the earthy smell of damp ground as you walk along a stream, and the taste of tangy wild blackberries you find growing in a sunny break in the trees. These and many other perceptions of your surroundings will continuously flow through your awareness, effortlessly and rapidly, as you interact with the environment. An even greater number of sensory neural responses to your environment will occur, although you will never become consciously aware of them. The same will be true for your friend, although what you each become aware of may differ based upon your individual interests and expectations as you take the walk.

By perceiving stimuli in the environment, humans and other species increase their likelihood of survival. Acquiring edible food, avoiding dangerous situations, and reproducing all become more likely if an organism can quickly perceive aspects of its current situation. Perceptions of our surroundings usually happen so rapidly and seem so complete that most people give them little thought. However, the multiple, simultaneous chains of events that occur in order for us to perceive our surroundings are extremely complex and still not completely understood. Further, what we perceive is not a perfect representation of the environment; it is simply one that is good enough in most ways most of the time for survival. Curiosity about perceptual processes has motivated philosophical discourse and scientific research for hundreds of years.

In addition to perception being explored simply to understand how it occurs (pure research approach), it is also studied for many practical, or applied, reasons. There are many ways by which perceptual processes can be damaged or not develop properly. The goal of many current researchers is to understand these processes and eventually develop ways to help individuals with perceptual problems. Another application of perception research involves human factors: the use of knowledge about how humans function (physically, perceptually, cognitively) to design safer and more efficient interactions between humans and objects such as machines or computers.

This chapter presents an overview of many aspects of perception. Historically and currently, how does the study of perception fit with other areas of psychology, especially those of sensation and cognition? What are some basic principles of perception? What are some of the challenges for and theories of perception? What have special populations of participants allowed us to learn about perception?

PERCEPTION AS A FIELD OF STUDY

How is perception distinguished from sensation (see Chapter 19)? There are mixed views on the issue. One simple distinction is that sensation is the neural response to a stimulus, and perception is our conscious awareness, organization, or recognition of that stimulus. Following this line of distinction, cognition is then the active, conscious manipulation of perceptions to encode them in memory, apply labels, and plan response actions. For much of the twentieth century, such distinctions allowed researchers to compartmentalize the topics of their research areas. Researchers who strictly followed such an approach often did not attempt to link their research questions and results to those from a different area.

Although in many ways this compartmentalization of topics made sense during several decades of the twentieth century, these areas of research haven't always been separated, and most current researchers no longer draw hard lines between these areas. For example, some very early and influential perceptual researchers such as Helmholtz studied anatomy of the senses, measured neural responses (of easily accessible touch neurons), studied conscious perceptions, and attempted to understand how they were all integrated.

So, why would distinctions between sensation, perception, and cognition make sense? First, when trying to understand any complex set of processes, a good strategy is to break them into basic components and study the components individually before investigating how they all interact. Many decades of sensation, perception, and cognition research built a foundation of basic knowledge within each of these areas.

Second, until relatively recently, the available methodological techniques limited the questions that researchers studied, resulting in different types of techniques and participants being used for the different areas. Traditional techniques used to measure the neural processes of sensation and the anatomical connections between neurons were invasive and led to the majority of basic sensation research being performed on animals. (See chapter 15 for a more complete description of neural recording methods.) Meanwhile, although animals can be studied invasively (there are still many ethical constraints, of course), animals are not able to describe their perceptions to researchers. Thus, when humans participated in perception research, different methodological techniques were necessary (e.g., Fechner's psychophysical methods, forced-choice procedures, multidimensional scaling, etc.; see Chapter 20 for more details). Therefore, much of the basic sensation and perception research used different populations of participants and different procedures. Linking the two sets of data is often difficult because of increasing differences between human and animal sensory mechanisms and brains as the focus moves from lower to higher levels of processing. The distinction between perception and sensation research is illustrated by Hochberg (1978), who gives very complete coverage to classic perception research and extremely minimal mention of any possible underlying physiological mechanisms. In those places where he does mention possible physiological mechanisms, he also explicitly mentions that they are not well understood, and that it is not clear that they actually contribute to the perceptual experiences measured with humans.

As a more concrete example of the traditional distinctions between sensation and perception research, imagine that you desire to pick up this book so you can read it. Several things would have to happen in order for you to process its location in depth so you could accurately and efficiently reach out to grab it. There are two types of depth cues you might use: monocular cues (e.g., occlusion, linear perspective, familiar size, motion parallax, etc.) or binocular cues (disparity, convergence, accommodation). The binocular cue of disparity is basically a measure of the differences in the relative location of items in the two eyes' views, and it is used for the process of stereopsis. Stereopsis is what allows people to see 3-D in the popular Magic Eye pictures or in random-dot stereograms (first created by Bela Julesz, 1961).

Of interest to sensation researchers, as you focused on the book to determine its depth, specialized cells in your primary visual cortex would modify their neural signals in response to the book's disparity. Some cells would be best activated when there is zero disparity (the object of interest is at the same depth as the point of focus), whereas others would prefer differing amounts of disparity either in front of or behind the point of fixation. Such signals have been studied in animals (e.g., Barlow, Blakemore, & Pettigrew, 1967) by using microelectrode recordings as a stimulus was systematically varied in depth.

Taking a perceptual rather than a sensation research focus, as you look at the book, you might also become aware of whether is it nearer or farther than some other object, say your notebook, which is also located on the table. If the two objects were relatively close to each other in depth, you would be able to simultaneously fuse the images from your two eyes into a single, 3-D image. If they were too far apart, the two eyes' views of only one of the objects could be fused into a single image; the other would be diplopic and appear as a double image. (If you pay close attention you can notice diplopia of nonfused objects; often our brains simply suppress the double image so we don't notice them.[1]) Ogle (1950) and Julesz (1971) systematically

[1]One easy way to demonstrate diplopic images to yourself is to hold up a single finger and put it ten inches or so in front of another object. Focus on the far object, but pay attention to your finger (wiggling it helps draw your attention, but remember not to focus directly on it). You should notice that your finger appears doubled, or diplopic. If you change your focus to your finger, you should be able to pay attention to the other object and notice that it appears diplopic. If you move your finger close to the far object, then neither should appear diplopic when you are focused on either one of them.

studied such relations by having humans report their subjective perceptions of fusion and diplopia as characteristics of the stimulus were systematically manipulated. Both wrote comprehensive books summarizing binocular vision, but neither attempted to link physiological and psychophysical data. This omission is quite reasonable for Ogle, given that the binocular neurons had not yet been discovered. However, it was also quite accepted that Julesz only briefly mentioned the physiological findings in a relatively cautious manner, and that he focused his writings on the perceptual aspects of binocular vision.

These examples of relatively pure sensation-oriented and perception-oriented research illustrate how important fundamental knowledge could be gained without requiring direct links between sensation and perception. There are a few counterexamples, where the level of processing was basic enough that even many decades ago researchers were able to fairly definitively link animal sensation data to human perceptual data. The most compelling of these examples is the almost perfect match between a human's measured sensitivity to different wavelengths of light under scotopic conditions (dim lighting when only rod receptors respond to light) and the light absorption of rhodopsin (the pigment in rods that responds to photons of light) across the different wavelengths. In most cases, however, differences in anatomy and how and what was measured made it difficult to precisely link the animal and human data. Despite these difficulties, as researchers gained understanding of the fundamentals of sensation and perception, and as technology allowed us to broaden the range of testable questions, more and more researchers have actively tried to understand perceptual processes in terms of underlying neural mechanisms. Clearer links between sensation and perception data have developed in several ways.

Because animals are unable to directly communicate their perceptions of a stimulus, researchers have limited ability to use animals to study how sensations (neural responses) are directly linked to perceptions, especially human perceptions. However, although they cannot talk, animals can behaviorally indicate some of their perceptions, at least for some tasks. For example, researchers can train some animals to make a behavioral response to indicate their perception of nearer versus farther objects. Then, as the researchers manipulate the location of the target object relative to other objects, they can not only measure the animal's neural responses but also record its behavioral responses and then attempt to link the two types of responses. Until the 1970s, such linking was done in two steps because animals had to be completely sedated in order to make accurate neural recordings. However, newer techniques allow microelectrode recordings in alert, behaving animals, so the two types of information could be simultaneously measured. Further, in many cases humans can be "trained" to make the same behavioral responses as the animals. Then, researchers can compare the animal and human responses and make a more accurate assumption of the neural responses in humans based on measures made using animals.

Another major approach that researchers are now using to infer the neural mechanisms of human perception (and other higher-level processes such as reading, memory, and emotions) is noninvasive brain activity recordings. There are several types of noninvasive brain activity technologies that have been developed (fMRI, PET, MEG, etc.; see Chapter 16). When using these technologies, participants perform a task while different measures of their brain activity are recorded, depending upon the type of technology (oxygen use, glucose use, electrical activity, respectively). The areas of the brain that are used to process a stimulus or perform a task show relatively increased use of oxygen and glucose, and those areas will also show changes in their levels and timing of the patterns of electrical activity. Both relatively basic and more complex perceptual processes have been studied. For example, these technologies have been used to study the neuronal basis of human contrast sensitivity (Boynton, Demb, Glover, & Heeger, 1999), to localize relatively distinct areas of the IT cortex that are used to process faces versus other objects (Kanwisher, McDermott, & Chun, 1997), and to localize different areas of the brain used to process different aspects of music perception (Leventin, 2006).

These noninvasive brain activity technologies all have relatively poor spatial resolution compared to microelectrode recordings. In other words, researchers can only localize a type of activity to within roughly a one-millimeter area of the brain. In comparison, invasive microelectrode recordings localize the specific activity being measured to a single, specific neuron. Thus, these two types of approaches (human brain activity measures and microelectrode recordings in animals) are still often difficult to link precisely. However, technology continuously brings animal (invasive) and human (noninvasive) methods closer together. For example, an advance in microelectrode recording technology allows researchers to record from arrays of neurons rather than a single neuron at a time. By recording from many neurons, researchers can better understand how the activity in one neuron influences other neurons, and they can better build a picture of the system processes as a whole. Multiple electrode recordings are also beginning to be used with humans in special cases—for example, individuals who are undergoing intracranial monitoring for the definition of their epileptogenic region (Ulbert, Halgren, Heit, & Karmos, 2001). Meanwhile, the spatial resolution of the noninvasive technologies is also improving. Eventually, researchers should be able to completely understand perception in terms of the activity of neural mechanisms, and the simple distinction between sensation and perception will become even less clear.

The division between perception and cognition is also blurring. For example, there are additional factors that would influence your reaching for a book, and these illustrate higher-level perceptual processes that interact

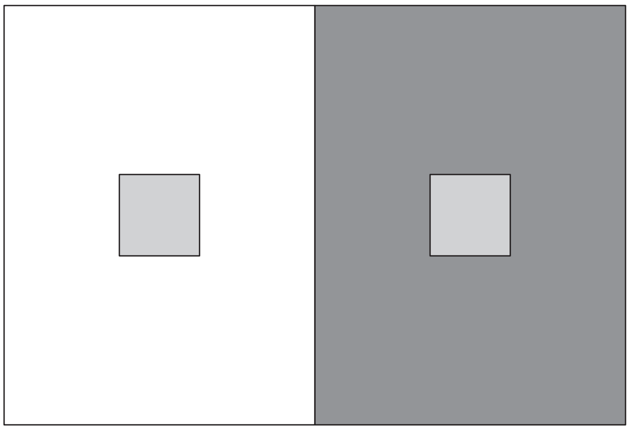

Figure 27.1 An example of simultaneous contrast, illustrating the influence of context on what we perceive. The inner squares are physically the same shade of gray, but most people will perceive that the gray square on the left surrounded by the white border is a darker shade of gray than the gray square on the right surrounded by the dark border.

with memory, which has traditionally been a research topic for cognitive psychologists. Although perceptual psychologists have long recognized that prior experience with stimuli can influence perceptions, they traditionally have not studied how the memories are formed, stored, or recalled. For example, based on past experience, you would have some memory of the size and weight of the book you desire to pick up. When you are determining its location in depth, your memory of these book characteristics will influence your determination of its location. Many perception researchers have investigated such interactions between stimulus-based characteristics and experience-based memories, but they have traditionally focused on the perceptual aspects rather than the cognitive ones. Related to the above book example, Ittleson and Kilpatrick (1952, as cited in Hochberg, 1978) manipulated the physical size of familiar playing cards and showed that participants mislocated them in depth (as long as there weren't other depth or object comparison cues available). The famous Ames room manipulates depth cues and relies on participants' previous experience with normal rooms (four walls that intersect at perpendicular angles). In the Ames room the walls are not perpendicular, but the visual perspective of the walls, windows, and other objects in the room have been created so that it appears "normal" when viewed through a monocular peephole. Viewers are fooled into perceiving well-known objects (e.g., a friend or family member) as growing or shrinking as they move through the room. As soon as the room is perceived correctly (by using a large viewing slot so both eyes can see the room and correctly interpret the unusual configuration of the walls and objects), the illusion is destroyed.

Similar to how new technologies have begun to better link sensation and perception, the new technologies are also allowing more links with topics that have traditionally been placed within the cognitive domain of research. Processes such as reading (typically considered a language-related process and within the domain of cognition) rely on perceptual input as well as language knowledge. Visual perception researchers have studied how factors such as the contrast, size, polarity, background texture, and spatial frequency of letters influence their readability (e.g., Legge, Rubin, & Luebker, 1987; Scharff & Ahumada, 2002; Solomon & Pelli, 1994). Others have focused on understanding dyslexia, and concluded that subtle changes in the magnocellular and parvocellular visual-processing streams can lead to dyslexia (e.g., Demb, Boynton, & Heeger, 1998). Greater understanding of memory processes (also traditionally within the cognitive domain) has indicated that memories for different characteristics of stimuli (visual, auditory, smell, etc.) rely on the brain areas originally used to process them. For example, if you visualize your room from memory, the pattern of brain activity observed will involve the visual-processing areas and highly resemble the pattern of activity that would be observed if you were actually viewing your room (Kosslyn et al., 1993). When individuals imagine hearing a song versus actually hearing that song, the patterns of brain activity are essentially identical (Levintin, 2006).

The blurring of the distinctions between sensation, perception, and cognition challenges current researchers, because they should now learn at least the basics of all these traditional areas. Although there are still some fundamental questions to be answered within each separate domain, understanding of what many consider to be the more interesting, complex questions will involve understanding interactive processes across multiple domains.

SOME BASIC PRINCIPLES OF PERCEPTION

Although the process of perception usually appears simple, complete, and accurate, it is not. This is true for each of the senses, as the following principles and examples will illustrate. These principles are being listed as if they were independent, but in reality they are interrelated.

1. *Regardless of the sense, sensory receptors only convert a restricted range of the possible stimuli into neural signals. Thus, we do not perceive many stimuli or stimuli characteristics that exist in the environment.* Further, because the design of the sensory receptors (and later perceptual systems) varies across species, different species perceive the environment differently. For example, humans can only perceive electromagnetic energy that has wavelengths between roughly 400nm and 700nm (the range that we refer to as "light"). We cannot perceive infrared or ultraviolet light, although other species are able to (e.g., some reptiles [infrared] and many birds, bees, and fish [ultraviolet]). Different species can perceive different ranges of auditory frequencies, with larger animals typically

perceiving lower ranges of frequency than smaller animals. Thus, humans cannot hear "dog whistles," but dogs (and other small animals) can. We can build technologies that allow us to perceive things that our systems don't naturally process. For example, a radio allows us to listen to our favorite song because it has been encoded into radio waves and transmitted to our location. But, even though it is physically present in the room, we can't hear it until we turn on the radio. It's important to note, however, that these technologies only allow us to perceive these other stimuli by converting them into stimuli that we can perceive naturally.

2. *The stimuli that are transformed into neural signals are modified by later neural processes. Therefore, our perceptions of those stimuli that are encoded are not exactly accurate representations of the physical stimuli.* We often enjoy such mismatches to reality through demonstrations using illusions. The shifts from reality can be due to genetically predisposed neural wiring (e.g., center-surround receptive fields that enhance the perception of edges for both vision and touch, or the motion detectors that allow us to perceive apparent movement, such as seen in the Phi Effect or in movies, which are a series of still frames). They can also be due to neural wiring configurations that have been modified based on experience (e.g., color perceptions can be shifted for up to two weeks after participants wear color filters for several days; Neitz, Carroll, & Yamauchi, 2002).
3. *What we perceive is highly dependent upon the surrounding context.* The physical stimulus doesn't change, just our perception of it. For example, after being in a hot tub, warm water feels cool, but after soaking an injured ankle in ice water, warm water feels hot. The color of a shirt looks different depending upon the color of the jacket surrounding it. Similarly, the inner squares in Figure 27.1 appear to be different shades of gray, but really they are identical. The taste of orange juice shifts dramatically (and unpleasantly) after brushing one's teeth. Many visual illusions rely on context manipulation, and by studying them we can better understand how stimuli interact and how our perceptual systems process information.
4. *Perception is more likely for changing stimuli.* For example, we tend to notice only the sounds from the baby monitor when the regular breathing is replaced by a cough or snort; we will be continually annoyed by the loose thread on our sleeve but not really notice the constant pressure of our watch; and we will finally notice a bird in a tree when it flies to a new branch. In all these cases, we could have heard, felt, or seen the unchanging stimulus if we had paid attention to it. Information about each of them was entering our sensory systems, but we didn't perceive them. Why? Sensory systems continuously process huge volumes of information. Presumably, we are prone to perceive those stimuli that are most useful for making good response choices and actions to promote survival. When a stimulus first appears, it will tend to be perceived because it represents a change from the previous configuration of the environment. It could be a threat or a stimulus that somehow should be acknowledged, such as a friend. However, once it has been noticed, if it does not change, the system reduces its neural response to that stimulus (the process of neural adaptation). This is a reasonable conservation of energy and metabolic resources, which allows us to focus our perceptions on items of greater interest.
5. *Perception can be altered by attention.* The principle regarding the influence of change alludes to such selective attention mechanisms. Attention can be drawn to stimuli by exogenous influences (such as the changes in stimuli noted previously, or any difference in color, loudness, pressure, etc., between stimuli), or attention can be placed on a stimulus due to endogenous factors (such as an individual's personal interests or expectations). When attention is focused on a specific stimulus, the neural response to that stimulus is increased even though nothing else about the stimulus has changed (e.g., Moran & Desimone, 1985). There are two types of perceptual demonstrations that illustrate well the influence of attention on perception.

   Change blindness demonstrations test our ability to perceive specific stimulus changes across different views or perspectives of a scene. The crucial factor is that the different perspectives are not temporally contiguous; there is some short break in our viewing of the scene (which can be less than 0.1 of a second), so that we have to hold the previous perspective in memory in order to make comparisons with the following perspective of the scene. This frequently happens in real life, such as when a large truck passes us and blocks our view of a scene, or when we turn our heads to look at something else and then turn back. Rensink (2002) systematically studied change blindness by using two views of a scene that altered on a computer screen with a blank screen interspersed between alterations. There was one obvious difference between the two views of the scene (e.g., the color of an object, the deletion of an object, a shift in placement of an object). He found that, unless participants were directly attending to the changing object, they had a very difficult time perceiving the change. Movie producers often rely on change blindness; across screen cuts there might be obvious changes in the objects in a room, or an actor's clothing, but unless you happen to be paying attention to that particular item, it is unlikely you will ever notice the change.

   Inattentional blindness demonstrations illustrate that we are often not aware of obvious, changing stimuli that are in the center of our field of view. For example, Simons and Chabris (1999) showed film clips of two teams playing basketball, instructing participants to attend to the ball and count how many times one of the teams passed the ball. The basketball game clip was modified so that a second clip was interleaved with it—one of a gorilla walking across the basketball court. When participants were attending to the ball, most of them did not report seeing the gorilla walking through the scene. You can view examples

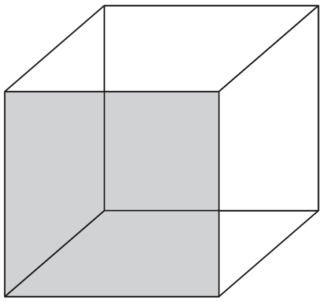

Figure 27.2 The Necker Cube.[1]

of this clip as well as other examples of change blindness and inattentional blindness on the Visual Cognition Lab of the University of Illinois Web site (http://viscog.beckman.uiuc.edu/djs_lab/demos.html).

6. *Perception is the result of both bottom-up and top-down processes.* Essentially all information individuals acquire of the world comes in through their senses (often referred to as *bottom-up processing*). This type of processing is what most people think about when they reflect on what they perceive—"I perceive my dog because my eyes see my dog" or "I feel the smooth texture of the glass because my finger is moving smoothly across it." Indeed, unless you are dreaming or hallucinating (or are part of *The Matrix*), your perceptions are dominated by external stimuli. However, the incoming sensory information is not the only influence on what you perceive. Of the millions of sensory signals being activated by environmental stimuli, only a small subset will enter conscious perception (those attended to). Of those, only a few are processed deeply enough to permanently change the brain (i.e., be encoded as a memory). These previously encoded experiences can influence your expectations and motivations, and in turn, influence the bottom-up sensory signals and the likelihood they will achieve conscious perception (often referred to as *top-down processing*). An increasing body of research has demonstrated top-down influences on the neural responses of relatively low-level sensory/perceptual processing areas of the brain (e.g., Delorme, Rousselet, Mace, & Fabre-Thorpe, 2004).

There are innumerable examples of top-down influences. If we are walking in the woods, and I tell you that there are a lot of spiders around, you might then perceive any soft movement on your skin as being a spider, at least for a second. Once you attend to it carefully and increase the bottom-up information that negates the spider interpretation, you will correctly perceive that it was just a leaf on a nearby bush. If I had not told you about spiders in the first place, you might not have even perceived the soft touch.

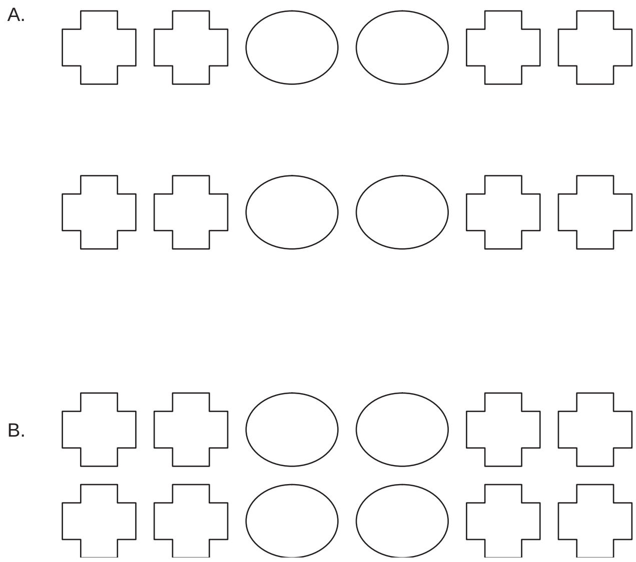

Figure 27.3 The Gestalt organization principles of proximity and similarity.[2]

This example shows how your perception can be fooled by top-down expectations. However, top-down influences are very useful. They can prime you to perceive something so that you can react to it more quickly. They can also help you perceive incomplete stimuli or interpret ambiguous stimuli, both of which happen in daily life (e.g., top-down processing helps us find objects in cluttered drawers or interpret a friend's handwriting). Top-down processes often lead two people to have different interpretations of the same event.

The above principles may imply that our perceptual systems are not very good.

There are things we cannot perceive, what we do perceive may not be exactly accurate, and our perceptions may change based on context, attention, or other top-down processes. However, our perceptual systems are for the most part amazingly accurate and have developed to maximize processing efficiency and promote our survival. If we perceived everything, we would be overwhelmed with information, most of which is not necessary for the task at hand. In the natural environment, the mismatches to reality will promote survival (e.g., by enhancing edges) more than threaten it, and the ability to use top-down as well as bottom-up processes allows for more flexible responses as a function of experience.

[1] This stimulus illustrates qualitative ambiguity because a viewer's interpretation of it can alternate between two different plausible configurations. In this case, both configurations are that of a cube, but the gray-shaded panel of the cube can be either the front-most surface of the cube, or the rear-most surface of the cube.

[2] In (A) most people tend to describe the stimuli as two rows made up of the shapes: plus, plus, oval, oval, plus, plus. The spacing between the two rows causes the shapes to group according to row rather than group by shape. In (B), the two rows are positioned so the vertical space between the shapes is equal to the horizontal space between the shapes. In this case, most people tend to describe the stimuli by grouping them by shape, i.e., a set of four pluses, then a set of four ovals, and then another set of four pluses.

Figure 27.4 A delayed perception using a fragmented figure.[4]

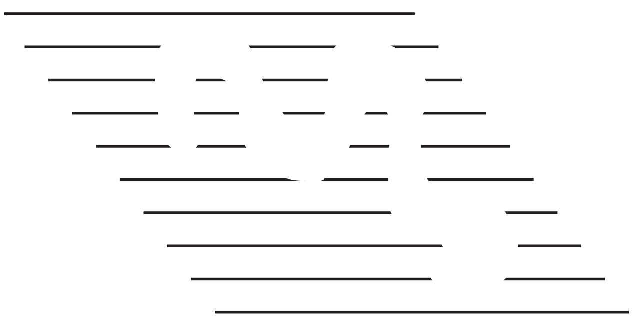

Figure 27.5 A subjective contour.[5]

In fact, our perceptual systems are able to process complex environmental surroundings more quickly and accurately than any computer. The stimulus information that activates our sensory receptors is often ambiguous and incomplete; however, our systems are still able to rapidly produce richly detailed and largely accurate perceptions. Do we learn to do this as we experience interactions with the environment, or are perceptual abilities controlled by innate mechanisms? The next subsection discusses some of the theories related to this question, along with some challenges a perceptual system must overcome when making the translation from stimulus energy to perception.

CHALLENGES FOR PERCEPTUAL SYSTEMS AND THEORIES OF PERCEPTION

There are many theories of perception, several of which have roots in age-old philosophical questions. (See Chapters 1 and 2 for additional summaries of some of the historical and philosophical influences in psychology.)

Regardless of the theory of perception, however, there are several challenges that any perceptual system or theory must handle in order to be considered complete. As already mentioned, our systems tend to quickly provide us with amazingly accurate perceptions, for the most part. Thus, researchers did not appreciate many of the challenges outlined in the following until they attempted to study the perceptual processes or create artificial perceptual systems. The following list of challenges was taken from Irvin Rock's (1983) book, *The Logic of Perception*. These challenges and examples tend to focus on vision, but they also hold true for the other senses (e.g., see Levintin, 2006, for multiple examples using music and other auditory stimuli).

1. *Ambiguity and Preference*. Rock recognized two types of ambiguity, dimensional and qualitative. Dimensional ambiguity refers to the fact that a single external stimulus (distal stimulus) can lead to multiple retinal images (proximal stimulus), depending upon the distal stimulus placement in our 3-D environment. For example, the retinal image of a business card could appear to be a perfect rectangle (front view), a trapezoid (due to an angled view with one side of the card closer to the eye so that its retinal image length is longer than the far side of the card), or a line (edge view). However, our perception of the business card will almost always be of a perfect rectangle. (Sometimes, if the viewing perspective is an uncommon one, the object may not be accurately recognized.) Qualitative ambiguity refers to the fact that two different distal stimuli can lead to the same proximal stimulus. In both cases we do not usually have any problems perceiving the correct distal stimulus because, somehow, we have a preference for a single, usually correct interpretation of the ambiguous proximal stimulus. There are exceptions to preference, but these generally occur for artificial stimuli such as the Necker Cube (see Figure 27.2), or other 2-D reversible figures.
2. *Constancy and Veridicality*. Even though proximal stimuli vary widely across different environmental conditions (lighting, placement, etc.), we tend to perceive a distal object as having color constancy, shape constancy, size constancy, and so on. In other words, we perceive the "true" distal stimulus, not one that varies as conditions change. Thus, somehow, perception becomes correlated with the distal rather than the proximal stimulus, even though the only direct information the perceptual system receives from the distal stimulus is the changing proximal stimulus.
3. *Contextual Effects*. As already explained, perceptions of distal stimuli are affected by neighboring stimuli, both in space and in time.
4. *Organization*. Our perceptions tend to be of discrete objects; however, the information in a proximal stimulus is simply a pattern of more or less electromagnetic energy of certain wavelengths; thus, it often does not clearly delineate object borders. For example, some objects may be partially occluded, and some "edges" may really be shadows or patterns on the surface of an object. With respect to organization, Rock highlights the Gestalt fundamentals of

[3] After several seconds of viewing, most people are able to decipher the phrase. (Give it a try. If you're stuck, read the following backwards: !!taerG er'uoY) After perceiving the "completed" figure, viewers will typically be able to more rapidly perceive it the next time it is viewed, even if a large amount of time has elapsed.

[4] Your perceptual system creates the impression of the white, worm-shaped figure that appears to be lying across the background of black horizontal lines. In reality, there is no continuous shape, only black lines that have strategic parts of them missing.

organization such as figure/ground, proximity, similarity, continuation, and so on. Figure 27.3 (A and B) shows how proximity and similarity can influence the perception of a group of objects.

5. *Enrichment and Completion*. Also recognized by the Gestaltists, sometimes what we perceive is more than what is actually available in the proximal stimulus. As their famous saying goes, "The whole is greater than the sum of its parts." This might happen because distal objects are partially occluded by other objects. Further, once an object is recognized, it becomes more than just the visual image of the object because perceptions such as functionality are included. Rock suggests that these completion and enrichment abilities are due to prior experience with the object being perceived. A second class of enrichment/completion occurs when we perceive something not present at all, such as subjective contours (see Figure 27.4) and apparent motion. Experience is not necessary to perceive this second class of enrichment/completion.
6. *Delayed Perceptions*. In most cases, perception occurs rapidly; however, in some special cases, a final perception can occur many seconds (or longer) after the proximal stimulus has started to be processed. Examples of this include stereopsis of random-dot stereograms (or the Magic-Eye pictures) and the processing of some fragmented pictures (see Figure 27.5). Usually, once a specific example of either type of stimulus is perceived, subsequent perceptions of that same stimulus are much more rapid due to top-down processing based on memories of the earlier experience.
7. *Perceptual Interdependencies*. Perceptions can change depending upon the sequence of stimuli, with a previous stimulus altering the perception of a subsequent stimulus. This is basically a temporal context effect, and it illustrates that it is not just information included in the proximal stimulus that influences how that stimulus is perceived.

In addition to the challenges outlined by Rock (1983), there seems to be at least one additional challenge that is currently getting increasing amounts of research attention: *Integration/Binding*. How do we end up with cohesive perceptions of stimuli? Even within a single sense, different aspects of a stimulus are processed by different populations of neurons. Further, many stimuli activate more than one sense. For example, as we eat, we experience the texture, the taste, and the smell of the food. When we vacuum a room, we see and hear the vacuum cleaner, and we can hear the sound change as we see and feel it go across different surfaces. How are those distinct perceptions united?

Now that the challenges for perception and perceptual theories have been outlined, what are some of the theories of perception that have been proposed? Regarding the question, "How do our perceptual abilities come to exist?" there have been two major theoretical perspectives, nativism and empiricism. Nativism proposes that perceptual abilities are innate rather than learned. Empiricism proposes that perceptual abilities are learned through meaningful interactions with the environment. As summarized by Gordon and Slater (1998), these two perspectives are rooted in the writings of several famous philosophers: John Locke and George Berkeley (supporting empiricist viewpoints), and Plato, Descartes, and Immanuel Kant (supporting nativism viewpoints). There is evidence supporting both perspectives. The nativism perspective is supported by both ethological studies (studies of animals in their natural environments) and some observations of newborn humans. For example, newly hatched herring gulls will peck the red spot on the beak of its parent in order to obtain food (Tinbergen & Perdeck, 1950, as cited in Gordon & Slater, 1998), and baby humans will preferentially turn their heads toward the smell of their mothers' breast odor (Porter & Winberg, 1999). In contrast, support is given to an empiricist perspective by observations that restricted sensory experience in young individuals will alter later perceptions. Kittens exposed to only vertical or horizontal orientations during the first several weeks after their eyes opened were impaired in their ability to perceive the nonexperienced orientation. If young human children have crossed eyes or a weak eye, then their stereoscopic vision will not develop normally (Moseley, Neufeld, & Fielder, 1998). Given the evidence for both perspectives, only a relatively small group of researchers have exclusively promoted one over the other. (See Gordon and Slater for a summary of both the philosophical background and the early works in the development of the nativism and empiricism perspectives.)

Under the domain of each of these broad perspectives, many theories of perception exist (not necessarily exclusively as either nativistic or empirical). For example, Rock (1983) outlined two categories of perceptual theories: stimulus theories and constructivist theories. Stimulus theories are bottom-up and require a relatively perfect correlation between the distal and proximal stimuli. Rock pointed out that this is not possible for low-level features of a stimulus (due to ambiguity, etc.), but it is possible that higher-level features could show better correlation. He uses J. J. Gibson's (1972) Theory of Direct Perception as an example of such a high-level-feature stimulus theory. Gordon and Slater (1998) categorize Gibson's theory as an example of nativism because Gibson assumed that invariants existed in the natural, richly complex environment, and that additional, constructive processes were not necessary for perception. According to Rock, although perception is obviously largely driven by the stimulus, pure stimulus theories should not be considered complete because they do not take into account the influence of experience, attention, and other top-down processes.

Constructivist theories suggest that the perceptual systems act upon the incoming bottom-up signal and alter the ultimate perception. Rock (1983) distinguishes two types of constructivist theories: the Spontaneous

Interaction Theory and cognitive theories. According to the Spontaneous Interaction Theory, as the stimulus is processed in the system there are interactions among components of the stimulus, between multiple stimuli, or between stimuli and more central representations. Rock credits the Gestalt psychologists with the development of the Spontaneous Interaction Theory (although they did not call it that). The spontaneous interactions are assumed to be part of bottom-up neural processes (although the specific neural processes have largely been unspecified). Because the Gestalt theory doesn't include many known influences of prior experience and memory, Rock also considered it to be incomplete. Gordon and Slater (1998) point out that the Gestalt psychologist Wolfgang Kohler used the Minimum Principle from physics to explain the Gestalt principles of organization; in turn, he supported an extreme nativist position (whereas other Gestalt psychologists such as Koffka and Wertheimer did not).

A cognitive theory, according to Rock (1983), incorporates rules, memories, and schemata in order to create an assumption about the perceived stimulus, but these interpretations are nonconscious processes. Rock believed this complex constructive process explains why perception is sometimes delayed, and that the cognitive theory also handles the other challenges for perception that he outlined. For example, constancy (the perception that objects appear to have the same shape, size, and color under differing perspectives and conditions) requires access to memory, which is not acknowledged in either the stimulus or spontaneous interaction theories.

Given its reliance on memory and prior experience, Rock's (1983) cognitive theory may seem to fit with an empiricist perspective, but he personally believed that there were meaningful distinctions between a cognitive theory and the empiricist perspective, at least as described by Helmholtz. According to Rock, prior experience alone does not include the use of logic, which can allow for inferences not directly based on prior experience. Similar to Rock's cognitive theory is Richard Gregory's theory (1980, as cited in Noe & Thompson, 2002) that Perceptions are hypotheses, which proposes that perceptions are the result of a scientific inference process based on the interaction between incoming sensory signals and prior knowledge. However, according to Gordon and Slater (1998), Gregory clearly supported the empiricist perspective.

Given the evidence of both innate and learned aspects to perception, it is not surprising that more current researchers have proposed theories that acknowledge both. For example, Gordon and Slater (1998) summarize Fodor's modularity approach, which includes perceptual input systems (largely innate in their processing) and more central systems that rely on learned experiences. Gordon and Slater also summarize Karmiloff-Smith's approach, which moves beyond simple modularity but still recognizes both innately predisposed perceptual processes and perceptual processes that rely on experience.

The above overview of theories of perception further supports the complexity of perception. If it were a simple process there would be little debate about how it comes to exist. The benefit of having a theory about any scientific topic is that it will give structure to frame new hypotheses and, in turn, promote new research studies. By having multiple theories, the same problem or question is approached from different perspectives, often leading to a greater expansion of understanding.

USING SPECIAL POPULATIONS TO STUDY PERCEPTION

The previous discussion of participants in perception studies distinguishes some relative benefits of using animals (fewer ethical constraints, which allows invasive measures) and humans (they have a human brain and can directly report their perceptual experiences). However, there are special populations of humans that have made important contributions to our understanding of perceptual processes. Persons with damage to perceptual processing areas of the brain and young humans (infant through early childhood) are two of these special populations. Each group presents unique challenges with respect to their use, but each also has provided valuable insight on perceptual processes that could not have been ethically obtained using normal adult humans.

The study of brain-damaged individuals parallels the use of lesion work in animals, but humans can give subjective reports of the changes in their perceptual abilities and perform high-level tasks that are difficult or impossible for animals. The major limitation of the use of brain-damaged humans is that the area and extent of damage is not precisely controlled. In most cases, strokes or head trauma will damage multiple brain areas, or not completely damage an area of interest. Thus, it is difficult to determine precisely which brain structures suffer the loss of specific perceptual (or other) abilities. However, by comparing several individuals with similar patterns of damage and by taking into account the results of precise lesion work in animals, some reasonably firm conclusions can be made.

Of special interest to perception researchers are studies of individuals with agnosia (the inability to remember or recognize some selective aspect of perception). Generally, agnosias are limited to the processing of information from one sense, with the other senses able to still process that aspect of perception. For example, a person with prosopagnosia (inability to process faces) can still recognize another individual from the person's voice or by feeling the person's face. Examples of additional documented visual agnosias include color agnosia (cerebral achromatopsia), motion agnosia (cerebral akinetopsia), and different forms of object agnosia. Some individuals can recognize objects but are not able to localize them in space or interact with them meaningfully using visual information. Affecting the sense of touch are astereognosia (inability to recognize objects by touch) and asomatagnosia (failure to recognize parts of one's own body). Careful study of such individuals has supported a modular perspective of perceptual processing, in that damage to specific populations or groups of neurons leads to specific deficits in perception.

The study of young humans is challenging in that, like animals, they cannot directly (through language) communicate their perceptual experiences, and they are limited in the range of tasks they are able to perform. However, through technology (e.g., noninvasive event-related potentials) and the development of many special techniques (e.g., preferential looking, the conditioned head turn, and the high-amplitude sucking procedure), major advances have been made in our understanding of perceptual development.

The use of human infant participants has been key to examining nativism versus empiricism. Aslin's (1981, as cited in Gordon and Slater, 1998) model of perceptual development summarizes many perceptual development studies and provides support for both philosophical perspectives. This model proposes three types of visual development at birth: undeveloped (which can be induced or remain undeveloped based on postnatal experience), partially developed (which can be facilitated, maintained, or show loss due to postnatal experience), and fully developed perceptual abilities (which can be maintained or show loss due to postnatal experience).

Perceptual development research has also advanced our understanding of and ability to help individuals with problems due to perceptual development disorders. As with the use of brain-damaged participants, perceptual development researchers have used human as well as animal participants because some controlled manipulations of development would not be ethical to perform on humans. For example, some human babies are born with strabismus (eye misalignment), which can lead to amblyopia (the loss of visual acuity in an ophthalmologically normal eye). The condition can be induced in animals, allowing investigation of the resultant impact on brain development (measured using invasive techniques), and allowing researchers to test possible treatments prior to using them with humans. However, of more direct interest to parents of such babies are the studies of other strabismic babies prior to and following treatments (by measuring the impact of treatment on observable behaviors and through the use of other noninvasive measures)

SUMMARY

Perception is a complex process, and although researchers have produced volumes of work in our endeavor to understand it, there is still much to be learned. Historically, visual perception research has dominated the research efforts of the other senses, due to both the better accessibility of the sensory and neural structures of vision and humans' tendency to put greater value on that sense. However, auditory research is also well advanced, and research on the other senses is progressing rapidly. Perception research is also interfacing with areas such as artificial intelligence and the study of consciousness. Greater understanding of the processes of perception will continue to impact daily living through the development of devices to aid individuals with perceptual problems (e.g., computer chips to replace damaged brain areas), the development of new technology (e.g., artificial noses to detect bombs or illegal drugs; multisensory virtual reality), and the design of everyday items (e.g., more readable highway signs and Web pages).

REFERENCES AND FURTHER READINGS

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology, 193*, 327–342.

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research, 39*, 257–269.

Delorme, A., Rousselet, G. A., Mace, M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research, 19*, 103–113.

Demb, J. B., Boynton, G. M., & Heeger, D. J. (1998). Functional magnetic resonance imaging of early visual pathways in dyslexia. *The Journal of Neuroscience, 18*(17), 6939–6951.

Enns, J. (2004). *The thinking eye, the seeing brain: Explorations in visual cognition*. New York: W. W. Norton & Company.

Gibson, J. J. (2002). A theory of direct visual perception. In A. Noe & E. Thompson (Eds.), *Vision and mind: Selected readings in the philosophy of perception* (pp. 77–89). Cambridge, MA: MIT Press.

Gordon, I., & Slater, A. (1998). Nativism and empiricism: The history of two ideas. In A. Slater (Ed.), *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 73–103). East Sussex, UK: Psychology Press Ltd.

Hochberg, J. (1978). *Perception* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Humphreys, G. W. (Ed.). (1999). *Case studies in the neuropsychology of vision*. East Sussex, UK: Psychology Press Ltd.

Julesz, B. (1961). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal, 39*, 1125–1162.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: The University of Chicago Press.

Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*(11), 4302–4311.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., et al. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience, 5*, 263–287.

Legge, G., Rubin, G., & Luebker, A. (1987). Psychophysics of reading. V. The role of contrast in normal vision. *Vision Research, 27*, 1165–1177.

Levintin, D. (2006). *This is your brain on music: The science of a human obsession*. New York: Dutton.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782–784.

Moseley, M. J., Neufeld, M., & Fielder, A. R. (1998). Abnormal visual development. In A. Slater (Ed.), *Perceptual development: visual, auditory, and speech perception in infancy* (pp. 51–65). East Sussex, UK: Psychology Press Ltd.

Neitz, J., Carroll, J., & Yamauchi, Y. (2002). Color perception is mediated by a plastic neural mechanism that is adjustable in adults. *Neuron, 35*, 783–792.

Noe, A., & Thompson, E. (Eds.). (2002). Perceptions as hypotheses. *Vision and mind: Selected readings in the philosophy of perception* (pp. 111–133). Cambridge, MA: MIT Press.

Ogle, K. (1950). *Researches in binocular vision*. Philadelphia: W. B. Saunders Company.

Porter, R., & Winberg, J. (1999). Unique salience of maternal breast odors for newborn infants. *Neuroscience and Biobehavioral Reviews, 23*, 439–449.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*, 245–277.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Scharff, L. F. V., & Ahumada, A. J., Jr. (2002). Predicting the readability of transparent text [Electronic version]. *Journal of Vision, 2*(9), 653–666.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*, 395–397.

Ulbert, I., Halgren, E., Heit, G., & Karmos, G. (2001). Multiple microelectrode-recording system for human intracortical applications. *Journal of Neuroscience Methods, 106*, 69–79.

PART V

Evolution and Behavior

28

Evolutionary Psychology

The Impact of Evolution on Human Behavior

Gordon G. Gallup, Jr., and Jeremy Atkinson
State University of New York, Albany

Daniel D. Moriarty
University of San Diego

Evolutionary psychology (EP) uses the theory of evolution to predict and explain human behavior. EP posits that the brain, just like any other organ (e.g., lungs, heart, etc.), has been shaped by natural selection and thus can be best understood from an evolutionary framework. This chapter outlines the theory of evolution and provides a number of concrete examples of EP using the latest research, complete with clinical and societal implications. Additionally, there is a separate section that addresses many common criticisms of EP and evolution.

EVOLUTION

Forget everything you thought you knew about evolution. Evolution does not create ideal creatures or abilities, nor is it based on the "survival of the fittest." Nobody is perfect and everybody dies; death and failure are an inevitable consequence of life. When it comes to evolution, it is not a question of whether you live, die, succeed, or fail; it is a question of whether you reproduce. Evolution is not about competition for food or other scarce resources but rather competition for genetic representation in subsequent generations. It is ultimately based on the perpetuation of genes, not individuals.

In its most rudimentary sense, evolution implies change. The simplest way to define evolution and represent such change is in terms of gradual changes in the composition of a gene pool over time. A gene pool is a hypothetical conglomerate of all of the genes being carried by all of the members of a particular species. In any particular gene pool, if individuals carrying certain configurations of genes leave relatively more (or relatively fewer) descendants than those carrying another configuration of genes, these gene frequencies will become progressively more (or less) prevalent in subsequent generations. In other words, the composition of the gene pool at any particular point in time is a direct reflection of the reproductive success of individuals in the immediately preceding generation. You could be the strongest, most disease resistant, most intelligent person alive, but if you don't reproduce your contribution to evolution is zero. This means that the decision to remain childless is equivalent to committing genetic suicide; the unique configuration of genes that make you who you are will not be included in the human gene pool in the future.

In order for anything to evolve, the probability of passing on its genes must be positively reinforced. An *adaptive trait*, therefore, is one that confers a reproductive advantage by increasing the probability of passing on one's genes, while a *maladaptive trait* is one that diminishes the likelihood of producing descendants. From this vantage point, *selection* can be defined as a correlation between the genetic basis for a trait and the probability of reproductive success. The size of the correlation determines the rate of selection, whereas the sign of the correlation determines the direction of selection. If the correlation is positive, there will be selection for the trait and it will become more prevalent; if the correlation is negative, there will be selection against the trait and it will eventually disappear. A few examples of adaptive traits that have been important to the evolution of modern humans are increased cranial capacity, bipedalism, and self-awareness.

Common Misconceptions of EP

Evolution works to improve the species. Evolution does not occur by design. Evolution occurs by selection, and the raw material for such selection consists of nothing more than random genetic accidents (mutations). Evolution does not work to improve the species, only the probability of reproductive success. The fact that over 99 percent of all the species that ever lived have gone extinct suggests that evolution works to the demise of most life forms.

Evolution has a purpose. Evolution lacks a conscious, intentional, or deliberate component. Each species is a by-product of a complex array of random genetic accidents. Differential reproduction is the key to evolution and what promotes and maintains various adaptations.

Evolution sustains human contentment. Evolution does not work to promote human happiness. The capacity for happiness, sadness, jealousy, envy, and all of the other human emotions are a by-product of evolution. For example, the ability to experience pain is an adaptive capacity. Organisms that experience pain learn to refrain from engaging in behaviors that are not in their biological best interests, such as putting one's hand on a hot stove. Likewise, happiness has been shaped by selection to occur when it is in our best interest to engage in a specific behavior.

EP tells "just-so stories." A common criticism of EP is that it generates an adaptive story that conveniently explains all the data amassed to date, and then claims that it has generated a perfect theory because all the data are explained! EP is based on *science*, not artful storytelling. Any hypotheses generated by EP must be parsimonious and testable, and the research results must be replicable by other scientists. If the results of research do not fit the theory, then either the theory must be modified to fit the evidence or the theory must be discarded. Furthermore, a theory must generate novel predictions. If a theory fails to fulfill any of these tenets, then the problem is not with science, but rather the theory.

Evolution is only a theory. The common usage of the word "theory" implies that there is no strong evidence in favor of the idea or concept, or that plausible alternatives exist. Evolution is based on a huge body of accumulating evidence from many scientific disciplines, including biology, geology, mathematics, genetics, and of course paleontology. There is no other scientific theory that can even begin to explain this body of evidence.

Evolutionists don't all agree on what evolution is (or, there are scientists who disagree with evolution). In any science, progress is based on controversy and disagreement. New ideas are tested, old ones rejected. EP is the same in this regard. However, one idea that is never disregarded or dismissed is evolutionary theory. Scientists disagree all the time about the nuances of evolutionary theory, but they never question evolution itself because there is so much diverse evidence in support of it.

Intelligent design disproves evolution. Intelligent design, try as it might, is simply not science. At some level it ultimately relies upon an unknown, unseen, and unproven agent that supposedly directs cosmic affairs. In the last analysis, intelligent design is based on faith, not fact. Evolution, on the other hand, is based entirely on scientific evidence.

EP is immoral. EP and evolution use science to describe the world around them, including human behavior. Just because EP finds that humans have violent impulses or that murder is a way to increase your fitness does not mean that this is the way society or humanity ought to be. EP can provide valuable insight into why people break the law, but it can't be used to condemn or condone breaking the law. Do not fall victim to the naturalistic fallacy: Simply because something evolved does not mean that it is right or justifiable. Nor does it mean that it cannot be changed.

EP is antifeminist. That males and females are biologically different is a fact. There is also a lot of evidence to show that males and females are psychologically different. EP does not attempt to justify such differences; rather, it is an attempt to study and better understand these differences. Males can be great parents and females great generals. Denying EP simply on ideological grounds is unscientific.

EP is antireligious. Simply because EP is based on science does not mean that there is an inherent disconnect between it and religion. Indeed, the mere existence of religion may be a by-product of evolution (Maser & Gallup, 1990). EP explains the adapted mind, and religion attempts to find the place of humanity in the cosmos. Perhaps the best way to resolve the issue is to ask whom one would rather see for a serious medical illness—a doctor or a priest? Many people would answer both, yet while both perform important roles, only one can legally practice medicine. Additionally, most major religions now concede that the tenets of evolution are not contrary to their religious teachings, including such a statement by Pope John Paul II.

Evolution and EP

In laymen's terms, EP looks for the adaptive benefits of specific human behaviors. Adaptiveness can be measured by determining the *genetic fitness* of the individual, or how many copies of an organism's genes are passed on to subsequent generations as a result of the behavior in question. In lieu of actual progeny, EP frequently employs substitute measurements such as the number of sexual partners a person has had, his or her attractiveness or status (as rated by members of either or both sexes), or his or her success in a particular endeavor or field (e.g., salary, GPA). However, as we will outline, there are myriad ways to measure fitness. EP also looks for and explains why some behaviors in a modern setting appear to be maladaptive.

Maladaptive Behaviors

Many people are afraid of spiders and snakes. Most people are far less afraid of speeding cars or driving above the posted speed limit. However, the chances of being killed in a traffic accident are much higher than of being killed or injured by a snake or a spider. Why do people show such “maladaptive” fearlessness of traffic? The answer is quite simple: Snakes and spiders posed significant problems to people during human evolutionary history; automobiles did not. Evolution is not a forward-looking process. Current adaptations were shaped by what has happened in the past. Novel situations, like weightlessness, crack-cocaine, or automobiles, would take thousands of years to shape specific evolutionary adaptations.

Costs and Benefits of Adaptations

Evolution never creates a perfectly adapted organism, even in a traditional environment. This is because each adaptation entails costs as well as benefits. For example, the benefits that derive from increased cranial capacity in humans had to be balanced against the increased risk of mortality during childbirth; bigger heads have a greater chance of getting stuck in the birth canal, which can result in the deaths of both mother and child. Even something purely behavioral, like a susceptibility to arachnophobia, incurs costs to the organism; it has to set aside space in the brain to maintain the behavior and spend energy to develop it while maturing. Add to this the chance that the behavior may never be used or that the behavior may develop incorrectly (many people often lose consciousness when suddenly confronted with a snake or spider) and the benefits do not always outweigh the costs. However, if, on average, the benefits exceed the costs, then the behavior in question will contribute toward genetic fitness and the trait will be maintained in the gene pool. The cost benefit ratio can be illustrated by examining your behaviors and their reproductive consequences. People who go to college have fewer children, on average, than people who do not, which would suggest that they have reduced genetic fitness. On the other hand, people who go to college tend to make more money, which increases the likelihood of the survival of each child, thereby increasing his or her chances of marrying and having children as well. Therefore, the reproductive costs of attending college can be offset by resulting economic advantages, which can increase your genetic fitness by increasing the chances that your progeny and their progeny will live to reproduce. During evolution, organisms that consistently behaved in ways in which the reproductive costs exceeded the benefits would have been selected against, and only those that behaved in ways in which the benefits more than compensated for the costs were able to prevail.

Inclusive Fitness and Cost/Benefit Ratios

Competition with other individuals for genetic representation in the next generation involves more than your reproductive success. Through a process known as *kin selection*, you also have a vested interest in the reproductive success of your relatives. This is because you share approximately 50 percent of your genes in common with members of your immediate family. Therefore, if you behave in ways that enable your brother or sister to produce two children that they would not otherwise have had, it would be equal to producing a child of your own. Likewise, because you share 25 percent of your genes in common with each of your nieces and nephews, any action that costs you one child but produces four children for one of your nieces would be neutral from a cost/benefit point of view. Taken together, this means that every human has a selfish, vested interest in aiding his relatives as long as the benefits outweigh the costs.

SEX DIFFERENCES

Contrary to popular opinion, there is no such thing as equality when it comes to sex and reproduction. Because men and women have very different reproductive best interests, the ways in which they maximize their fitness can be very different; each sex faces different constraints in reproduction and thus has different costs and benefits associated with the same action or choice. In order to understand these cost/benefit differences, the three underlying biological dimensions that distinguish males from females will be examined.

Genetic Assurance

Females have a significant advantage when it comes to the question of sharing genes in common with their offspring. Maternity is always certain. Mothers have an ironclad guarantee of sharing 50 percent of their genes in common with each of their children. Paternity, on the other hand, is rarely certain. Because of the possibility of rape or female infidelity, males have to contend with the prospect of being cuckolded (i.e., being duped into caring for children sired by other males). The incidence of nonpaternity can be substantial. Among males with low paternity confidence (i.e., those that believe they have reason to question the paternity of their ostensible children), the worldwide incidence of nonpaternity is as high as 30 percent (Anderson, 2006).

Parental Investment

Another important underlying biological dimension that distinguishes males from females is the issue of investment in children. Whereas the benefits that accrue to your fitness as a consequence of producing a child are the same whether you are a male or a female, the costs are not equally distributed. Females pay the lion's share of the parental investment tab. There is no such thing as a "Dutch treat" when it comes to reproduction. Females are the ones who get pregnant, experience childbirth, and serve as the primary caretaker for the child. This latter point is particularly true of mammals, for which the mother's milk is the primary source of nutrition for the offspring. Among many sexually reproducing species, the male's role in reproduction is to simply serve as a source of complementary gametes. Whereas the male's role in reproduction often focuses on insemination, for females insemination is the mere beginning of the reproductive process, a process that can take years of investment coupled with extended periods of caring and commitment by the female until the child reaches adolescence and can begin to fend for itself.

Reproductive Potential

Sperm are plentiful, but eggs are scarce. A single human ejaculate can contain as many as 250 million or more sperm. Every male produces millions of gametes on a daily basis. In stark contrast, females are born with all of the eggs they will ever have, and ovulation usually involves the release of a single ovum. Women typically ovulate only once during a normal menstrual cycle.

When a woman gets pregnant, it puts her capacity for further reproduction on hold. Pregnancy produces hormonal changes that serve to inhibit further ovulation, and as a consequence pregnancy prevents reimpregnation. Moreover, breast-feeding also leads to hormonal changes that serve to inhibit ovulation (a result called *lactational anovulation*), and during human evolutionary history mothers probably breast-fed infants for two to three years or longer.

Not only do males have an almost (theoretically) unlimited capacity to reproduce, but they also remain reproductively viable much longer than females. Assuming the typical male begins to produce viable semen at 15 years of age, many males continue to be capable of having children well into their 60s and even 70s. Thus a male who lives to be 75 years old may be reproductively viable for up to 80 percent of his life. If given access to enough reproductively viable women, a man could produce hundreds of offspring in a lifetime. Females, however, have to contend with a much more limited reproductive capacity and a truncated reproductive life span. As a consequence of menopause, most females lose the capacity to reproduce long before they die. Assuming that females begin to ovulate at about age 15 and menopause can occur in the late 30s and early 40s, females are capable of having children for only about 35 percent of their lives.

Reproduction is very costly for females. Unlike males, females can ill afford to make many reproductive mistakes. Thus females have a strong interest in the other 50 percent of their child's genes that derive from the father. Evolutionary theory would predict that women should be careful comparison shoppers when it comes to mate selection. Whether a woman's children can compete effectively for a limited number of reproductive opportunities in subsequent generations and perpetuate her genes would be influenced not only by the genes that she gives them but also by the characteristics they get from the father. Clearly, females who made judicious mate choices would have left more descendants than those who were indiscriminant about whom they had sex with.

On the other hand, because paternity is uncertain and males have an enormous capacity to reproduce, men have been selected to be opportunistic maters. This point is made in a compelling way by the results of the following classic study (Clark & Hatfield, 1989). Both male and female college students were approached individually by an attractive member of the opposite sex (a confederate of the experimenter), who then made one of three graded hypothetical invitations. Students were invited to (a) date the person, (b) go back to the person's apartment, or (c) have sex with the person. Whereas 50 percent of both the male and female students agreed to the date, only 6 percent of the females agreed to return to the person's apartment, and none accepted the invitation to have sex with someone they did not know. In stark contrast, more than 67 percent of the males agreed to go back to the person's apartment, and 75 percent agreed to have sex.

MENTAL MODULES

The next four sections examine a specific adaptive behavior. The section on incest avoidance introduces the concept of a mental module—an adaptive behavior that has evolved to solve a specific problem, task, or environmental feature encountered by most members of a species or sex. A mental module can evolve if the same situation is encountered over many generations; the preceding example of our species' susceptibility to arachnophobia is a good example of a mental module.

Incest Avoidance

A good example of the ability of evolutionary psychology to explain existing data, discount alternative theories, as well as generate and test novel hypotheses can be found by examining incest avoidance. There is ample evidence to suggest that the incest taboo is the result of an innate, evolved mental module, as opposed to a result of cultural values or society (e.g., the environment).

To begin with, there are large fitness costs associated with close-relative mating (i.e., between full siblings or offspring and parents). Much of this cost is due to inbreeding depression. All individuals carry a number of deleterious mutations that are masked because they also have a functional copy of the gene. Most of these mutations are inherited from parents and grandparents. Thus, if you mate with a close relative, you dramatically increase the chances of receiving two defective copies of the same gene. Researchers have documented inbreeding depression in many species. In humans, hemophilia, infertility, and microcephaly have all been traced to inbreeding. Additionally, a number of genes such as those that regulate the immune system maximize an organism's fitness when they are all different (i.e., the organism is heterozygous for the genes in question). Acquiring a novel set of these genes requires mating with a nonrelative (Potts, Manning, & Wakeland, 1991). As with inbreeding depression, mating with a close relative would increase your chances of receiving two copies of the same gene, and lead to a depressed immune system. In humans, matings between close relatives result in at least twice the rate of mortality and morbidity compared to control populations (Adams & Neel 1967; Seemanova, 1971).

Second, incest is a universal taboo (Brown, 1991) and is rarely practiced in human cultures. In almost every society, both past and present, there are very stringent rules prohibiting either sexual relationships or marriages between closely related individuals (Bixler, 1981). Furthermore, in the few societies that do condone a form of incest, the incestuous relationships have been almost entirely limited to the royalty or nobility (Bixler, 1982a, 1982b). Likewise, in the multitude of animal species examined to date, close-relative incestuous matings generally comprise less than 2 percent of all matings (Harvey & Ralls, 1996)

Third, there is also a large body of evidence that incest avoidance is triggered by mechanisms activated by cohabitation at a young age; for example, if you lived together with someone as a child, you will not be sexually interested in them as an adult. In other words, the fact that familiarity breeds contempt may be the result of selection against inbreeding. Some of the best evidence in support of this comes from a detailed study of the marriage patterns of former members of a *kibbutz* (a communal housing complex for children in Israel). Sherper (1971a, 1971b, 1983) found that out of 2,769 subjects, not a single one married a peer with whom they had been continuously raised for the first six years of life. Furthermore, detailed interviews with 65 individuals found only one incidence of postpubertal heterosexual activity among them, in which the person in question joined the kibbutz group at age 10.

Another line of evidence comes from Sim Pua marriages in China. These marriages, known as "minor form" marriages, involve the transfer of the bride to the groom's home at a young age, often as young as three years old. Wolf (1966, 1968, 1970, 1995) compared 132 Sim Pua marriages to 172 traditional marriages and found higher rates of divorce, separation, and adultery in Sim Pua marriages. Most interestingly, Sim Pua marriages produced 30 percent fewer children than regular marriages. These results are consistent with the suggestion that there is an innate incest taboo mechanism operating in humans mediated by early childhood rearing experiences.

There are a number of implications that arise from this analysis. Because the formation of the innate incest taboo requires cohabitation at young ages, families that don't expose siblings to one another as young children risk the possibility of incestuous relationships between the siblings later in life. Likewise, there is a risk of "quasi"-incestuous relationships between stepsiblings if they were not raised together as young children.

What we know about inbreeding depression demonstrates that the biblical account of human origins based on the story of Adam and Eve is untenable. This account implies that human origins were inherently incestuous. After Adam and Eve reproduced and had children, the only way to get to the next generation would have been for their children to have sex with one another, or for the children to have sex with their parents. Because there simply are no other alternative means by which their genes could be propagated, the resulting inbreeding depression would have had disastrous consequences for the human species.

SIGNALING AND MEASURING GENETIC FITNESS

The next three sections examine how humans detect genetic fitness in others and signal their own genetic fitness. There are many mental modules that function to determine the genetic fitness of other people. We selected these modules to focus on *honest signals of fitness*, signals that are truly indicative of the ability of an individual to pass on his or her genes. Each remaining section title is an example of an honest signal of genetic fitness. However, humans have also been selected to subvert honest signal detection modules. This is frequently accomplished by employing *dishonest signals of fitness*, signals that mimic honest signals in order to fool other people into thinking you (or your kin) are more genetically fit than you really are.

Paternal Resemblance

A common refrain among mothers is that all of their children look like their father, not them. In order to appreciate why this is the case, recall that women have at least one substantial reproductive advantage over men: Females can be virtually certain of sharing 50 percent of their genes in common with each of their children. However, as a consequence of rape or female infidelity, males have to contend with paternal uncertainty and being cuckolded (investing in offspring sired by other males).

Paternal resemblance, or the degree to which children look like their father, is one hypothesized phenotypic feature that fathers could use to gauge relatedness to their purported offspring. By presenting people with pictures of infants and asking them to identify their parents out of an array of adult photos, Christenfeld and Hill (1995) found that participants were better at matching children's pictures to their father than to their mother, although researchers have not always replicated this finding (Brédart & French, 1999). However, another study that had third parties *directly* compare the degree of resemblance of children to both parents found that after the age of three, boys resemble their fathers more than their mothers (Alvergne, Faurie, & Raymond, in press). If children show a paternal resemblance bias it may be a consequence of the fact that during human evolutionary history males invested preferentially in children who looked like them. Because of the enormous costs of cuckoldry, males who were indifferent to the question of paternity would have been at a significant disadvantage compared to males who cared about the paternity of their mate's children.

To test this hypothesis, Platek, Burch, Panyavin, Wasserman, and Gallup (2002) combined individual pictures of college students with one of several infant pictures using a computerized morphing program. Students were then presented with an array of five baby pictures that had been morphed with different adult faces, including the one that had been combined with their own face. When asked to make hypothetical investment decisions about the babies in these pictures (e.g., which one of these children would you adopt, or spend $50 on?), men, in direct contrast to women, showed a clear preference for investing in children who shared some of their own facial features. Another study that interviewed British men found paternal investment increased the more a father believed his children resembled him, and this effect was more pronounced for divorced fathers (Apicella & Marlowe, 2003). These results suggest that men do in fact invest preferentially in children as a function of their shared resemblance.

The existence of a paternal resemblance investment bias has been replicated in several subsequent studies (e.g., Anderson, Kaplan, & Lancaster, 2007; Platek, Critton, et al., 2003), and has been extended to show that male (but not female) brains are uniquely activated by pictures of children they resemble (Platek, Rainesa, et al., 2003).

There is even evidence that males modulate the importance of specific physical features based on how relevant these cues are to their specific phenotype in determining paternal certainty. Laeng, Mathisen, and Johnsen (2007) found that blue-eyed men (but not brown-eyed men or women of either eye color) found pictures of women with blue eyes more attractive than the *same pictures* featuring the *same women* with brown eyes. These findings remained statistically significant even after controlling for the color of either parent's eyes. Researchers hypothesized that blue-eyed men showed this difference because a blue-eyed man who mated with a blue-eyed woman would have only blue-eyed offspring. Thus, a child who exhibited a different eye color would obviously not be his. It is interesting to note that brown-eyed men, who would not gain any additional information by preferring blue-eyed women, exhibited no preference for eye color. The same study also found evidence for assortative mating: Blue-eyed men preferentially choose light-eyed partners.

Given the importance of paternal certainty, it is no wonder that there is a substantial body of evidence that mothers and the matriline comment more frequently on an infant's purported paternal resemblance than on its maternal resemblance (Daly & Wilson, 1982; Regalski & Gaulin, 1993). For example, McLain, Settersa, Moultona, and Pratt (1999) found that mothers of one- to three-day-old newborns made more comments about paternal than maternal resemblance, particularly when the father was present.

Preferential investment based on paternal resemblance has some interesting practical implications: In the case of adoption, by matching features of the adopted children (e.g., hair color, eye color) with features of the adoptive father there is reason to believe you might be able to promote better adoption outcomes. Indeed, these findings would even suggest that the risk of child abuse might be inversely proportional to the degree of paternal resemblance. There is some evidence to support this hypothesis: Burch and Gallup (2000) found that for men enrolled in a domestic violence treatment program, paternal resemblance was positively correlated with the self-reported quality of the men's relationships with their children and inversely proportional to the severity of injuries suffered by their spouses. In light of this study and the findings of Laeng, Mathisen, and Johnsen (in press), social service agencies might do well to target instances of obvious failures to show paternal resemblance for special intervention.

Both fathers and mothers act in ways that suggest paternal resemblance is very important. As with all evolved behaviors, the mental processes involved in assessing paternal resemblance need not be explicitly known to the individual. However, as both fathers and mothers actively and easily assess paternal resemblance, and modulate their behavior accordingly, it is quite probable that paternal resemblance is an example of an evolutionary adaptive behavior that is, at some level, consciously processed.

Fluctuating Asymmetry

Not everyone has features on one side of their body that perfectly mirror the features on the other side. For example, the length of the ring finger, the diameter of the wrist, or the length of the ear lobe may be slightly different from one side to the other. Researchers believe that random morphological deviations from perfect bilateral symmetry, known as *fluctuating asymmetry* (FA), are a consequence of stress-induced developmental instabilities that occur primarily during prenatal growth. Although the body is programmed to develop symmetrically, it is thought that environmental stressors (e.g., diseases, parasites, malnutrition) disrupt the normal developmental program. Individual differences in the magnitude of FA are taken as evidence for underlying genetic differences in the ability to buffer against such stressors. An individual with high FA thus would have either an increased exposure or a compromised ability to mitigate the effects of environmental stressors. Whatever the cause, it would be an indicator of decreased fitness. Indeed, there is a wealth of evidence that low FA individuals have increased fitness.

Researchers have correlated elevated FA with increased vulnerability to disease as well as higher levels of morbidity (Livshits & Kobyliansky, 1991), whereas low FA has been correlated with increased genetic, physical, and mental health (Thornhill & Møller, 1997) as well as increased fertility (Thornhill & Gangestad, 1994) and athleticism. Additionally, intelligence appears to be related to fluctuations in FA, with low FA males scoring significantly higher on standardized intelligence tests (Furlow, Gangestad, & Thornhill,1997).

In light of this evidence, it should not be surprising that people preferentially respond to low FA (high fitness) individuals. For example, men and women with low FA are consistently judged as being more attractive than those with high FA (Thornhill & Gangestad, 1994). Low FA men have more sexual partners, and they are more likely to engage in instances of sexual infidelity (Gangestad & Thornill, 1997). Women who are in committed sexual relationships with low FA men experience more frequent orgasms than those with high FA partners (Thornhill, Gangestad, & Comer, 1995). Even dancing ability appears to be related to FA—low FA dancers were rated as better dancers than their high FA counterparts, particularly by low FA individuals (Brown et al., 2005).

Low FA is even related to body odor. Thornhill and Gangestad (1999a) had a sample of male volunteers agree to wear freshly laundered T-shirts to bed and return the shirts to the laboratory the next day. The shirts were placed in coded plastic bags and each of the donors was measured for deviations from bilateral symmetry across a variety of physical traits. Women, who did not know the source of the shirts or the identity of the men who had worn them, smelled the shirts and rated the extent to which they found the odor of the different T-shirts to be attractive. Women who were not taking birth control pills and were in the ovulatory phase of their menstrual cycle rated the shirts worn by low FA men as smelling more attractive than those worn by high FA men. In other words, fertile women found the body odor of low FA men more attractive.

People also find that low FA individuals sound more attractive. Hughes, Harrison, and Gallup (2002) measured FA in male and female college students and had them individually count from 1 to 10 while their voices were being recorded. These recordings were then rated for attractiveness by a sample of men and women who did not know and had not seen the people whose voices had been recorded. Hughes et al. discovered that the voices of people with low FA were rated as being more attractive than those of people with high FA.

Mate choice is an important dimension of fitness. When it comes to competing for genetic representation in subsequent generations it is important to pick a mate who has high-quality genes. Offspring with high-quality parents have a better chance of showing traits that will bode well for their own reproductive success, which in turn will serve to propagate their parents' genes. The above evidence strongly suggests that humans have been selected to prefer a variety of subtle features in mating partners that are correlated with low FA in order to maximize their genetic fitness.

Facial Attractiveness

Faces are important. We recognize people based on their facial features, and the allure of those features plays an important role in interpersonal attraction. Many people believe that facial attractiveness is a social construction, driven by learned, arbitrary sociocultural influences. However, data have begun to emerge that challenge this position (for an earlier review of this evidence see Thornhill & Gangestad, 1999b). Not only is there relatively good consensus among people as to which faces are attractive, but attractive faces also appear to be a cross-cultural universal. That is, faces that people in Asia find attractive are also rated as attractive by people in North America, and vice versa. Ratings of facial attractiveness are consistent not only across cultures and ethnic groupsbut across sexes, sexual orientations, and ages as well. Coupled with the fact that even neonates spend more time looking at faces that we judge attractive, these data suggest that facial attractiveness may have a biological component.

There is growing evidence that individual differences in facial attractiveness signal important underlying biological properties. People with attractive faces are healthier (Shackelford & Larsen, 1999), and they tend to have lower FA (Grammer & Thornhill, 1994). Contrary to the social constructionist position, we may have been selected during human evolutionary history to find certain faces attractive because they were associated with honest signals of fitness. In other words, there may have been selective benefits to mating with people based on their facial features.

Several recent studies provide strong support for this position. In one recent study, Soler et al. (2003) took facial photographs and collected semen samples from 66 male college students. Semen quality for each man was determined by measuring sperm count, sperm motility, and sperm morphology. Then, a large sample of women were shown the photos of these men and asked to rate them for attractiveness. As with low FA males, ratings of facial attractiveness were significantly correlated with semen assay results; men with attractive faces tended to have higher-quality semen.

In another study, Henderson and Anglin (2003) selected 50 facial photographs, equally representing men and women, from old high school yearbooks. Using a public database that included the dates of birth and death for the people in these yearbooks, longevity data were calculated to the nearest month. Male and female college students were instructed to rate each of the yearbook photos for attractiveness and health. Ratings of facial attractiveness were highly correlated with ratings of health. Moreover, ratings of facial attractiveness of the high school photos were significantly correlated with longevity. Both men and women with attractive faces tended to live longer.

SUMMARY

EP has the ability to explain and predict many facets of human behavior. Based on the principles of evolution, it follows that behavior is also subject to natural selection, just like physical features. EP can be used to suggest new clinical and social treatment programs as well as to examine existing regimes for their efficacy. In recent years, EP has been gaining acceptance in all fields of mainstream psychology. Some people believe it has the potential to unite these disparate fields under a single and overarching theoretical framework.

REFERENCES AND FURTHER READINGS

Adams, M. S., & Neel, J. V. (1967). Children of incest. *Pediatrics, 40,* 55–62.

Alvergne, A., Faurie, C., & Raymond, M. (in press). Differential facial resemblance of young children to their parents: Who do children look like more? *Evolution and Human Behavior*.

Anderson, K. G. (2006). How well does paternity confidence match actual paternity? Evidence from worldwide nonpaternity rates. *Current Anthropology, 48,* 511–518.

Anderson, K. G., Kaplan, H., & Lancaster, J. B. (2007). Confidence of paternity, divorce, and investment in children by Albuquerque men. *Evolution and Human Behavior, 28*(1), 1–10.

Apicella, C. L., & Marlowe, F. W. (2004). Perceived mate fidelity and paternal resemblance predict men's investment in children. *Evolution and Human Behavior, 25*(6), 371–378.

Bixler, R. H. (1981). Comparative ethology of incest avoidance. In R. Fox (Ed.), *Biological anthropology* (pp. 37–67). London: Malaby Press.

Bixler, R. H (1982a). Comment on the incidence and purpose of royal sibling incest. *American Ethnologist, 9,* 580–582.

Bixler, R. H. (1982b). On genes, environment and behaviour. *Current Anthropology, 23,* 581–582.

Brédart, S., & French, R. M. (1999). Do babies resemble their fathers more than their mothers? A failure to replicate Christenfeld and Hill (1995). *Evolution and Human Behavior, 2,* 129–135.

Brown, D. E. (1991). *Human universals.* Philadelphia: Temple University Press.

Brown, W. M., Cronk, L., Grochow, K., Jacobson, A., Liu, C. K., Popovic, Z., et al. (2005). Dance reveals symmetry especially in young men. *Nature, 438*(22), 148–150.

Burch, R. L., & Gallup, G., Jr. (2000). Perceptions of paternal resemblance predict family violence. *Evolution and Human Behavior, 21*(6), 429–435.

Christenfeld, N., & Hill, E. (1995). Whose baby are you? *Nature, 378,* 669.

Clark, R. D., & Hatfield, E. (1989). Gender differences in receptivity to sexual offers. *Journal of Psychology and Human Sexuality, 2,* 39–55.

Daly, M., & Wilson, M. (1982). Whom are newborn babies said to resemble? *Ethology and Sociobiology, 3,* 69–78.

Furlow, A. P., Gangestad, S. W., & Thornhill, R. (1997). Fluctuating asymmetry and human IQ. *Proceedings of the Royal Society, London: Biological Sciences, 264,* 823–830.

Gangestad, S. W., & Thornhill, R. (1977). The evolutionary psychology of extrapair sex: The role of fluctuating asymmetry. *Evolution and Human Behavior, 18,* 69–88.

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology, 108*(3), 233–242.

Harvey, P. H., & Ralls, K. (1986). Do animals avoid incest? *Nature, 320,* 575–576.

Henderson, J. J. A., & Anglin, J. M. (2003). Facial attractiveness predicts longevity. *Evolution and Human Behavior, 24,* 351–356.

Hughes, S. M., Harrison, M. A., Gallup, G., Jr. (2002). The sound of symmetry: Voice as a marker of developmental instability. *Evolution and Human Behavior, 23,* 173–180.

Laeng, B., Mathisen, R., & Johnsen, J. A. (in press). Why do blue-eyed men prefer women with the same eye color? *Behavioral Ecology and Sociobiology.*

Livshits, G., & Kobyliansky, E. (1991). Fluctuating asymmetry as a possible measure of developmental homeostasis in humans: A review. *Human Biology, 63,* 441–466.

Maser, J. D., & Gallup, G. G., Jr. (1990). Theism as a byproduct of natural selection. *Journal of Religion, 70,* 515–532.

McLain, D. K., Settersa, D., Moultona, M. P., Pratt, A. E. (1999). Ascription of resemblance of newborns by parents and nonrelatives. *Evolution and Human Behavior, 21*(1), 11–23.

Møller, A. P., Soler, M., & Thornhill, R. (1995). Breast asymmetry, sexual selection, human reproductive success. *Evolution and Human Behavior, 16,* 207–219.

Platek, S. M., Burch, R. L., Panyavin, I. S., Wasserman, B. H., Gallup, G., Jr. (2002). Reactions to children's faces: Resemblance affects males more than females. *Evolution and Human Behavior, 23*(3), 159–166.

Platek, S. M., Critton, S. R., Burch, R. L., Frederick, D. A., Myers, T. E., & Gallup, G., Jr. (2003). How much resemblance is enough? Sex difference in reactions to resemblance, but not the ability to detect resemblance. *Evolution and Human Behavior, 24*(2), 81–87.

Platek, S. M., Rainesa, D. M., Gallup, G., Jr., Mohamed, F. B., Thomson, J. W., Myers, T. E., et al. (2003). Reactions to children's faces: Males are more affected by resemblance than females are, and so are their brains. *Evolution and Human Behavior, 25,* 394–405.

Potts, W. K., Manning, C. J., & Wakeland, E. K. (1991). Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature, 352,* 619–621.

Regalski, J. M., & Gaulin, S. J. C. (1993). Whom are Mexican infants said to resemble? Monitoring and fostering paternal confidence in the Yucatan. *Ethology and Sociobiology, 14,* 97–113.

Seemanova, E. (1971). A study of children of incestuous matings. *Human Heredity, 21,* 108–128.

Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior, 20,* 71–76.

Sherper, J. (1971a). Mate selection among second-generation kibbutz adolescents and adults: Incest avoidance and negative imprinting. *Archives of Sexual Behavior, 1,* 293–307.

Sherper, J. (1971b). *Self-imposed incest avoidance and exogamy in second-generation kibbutz adults.* Unpublished doctoral dissertation, Rutgers University, Rutgers, NJ.

Sherper, J. (1983). *Incest: A biosocial view.* New York: Academic Press.

Shoup, M. L., Pipitone, N. R., Scheyer, D., & Gallup, G. G., Jr. (2007, April). *Ratings of facial attractiveness predict grade point average in males, but not females.* Paper presented at the meeting of the Northeastern Evolutionary Psychology Society, New Paltz, NY.

Soler, C., Nunez, M., Gutierrez, R., Nunez, J., Medina, P., Sancho, M., et al. (2003). Facial attractiveness in men provides clues to semen quality. *Evolution and Human Behavior, 24,* 199–207.

Thornhill, R., & Gangestad, S. W. (1994). Human fluctuating asymmetry and sexual behavior. *Psychological Science, 5,* 297–302.

Thornhill, R., & Gangestad, S. W. (1999a). The scent of symmetry: A human pheromone that signals fitness? *Evolution and Human Behavior, 20,* 175–201.

Thornhill, R., & Gangestad, S. W. (1999b). Facial attractiveness. *Trends in Cognitive Sciences, 3,* 452–460.

Thornhill, R., Gangestad, S., & Comer, R. (1995). Human female orgasm and mate fluctuating asymmetry. *Animal Behaviour, 50,* 1601–1615.

Thornhill, R., & Møller, A. P. (1997). Developmental stability, disease and medicine. *Biological Reviews, 72,* 497–548

Wolf, A. P. (1966). Childhood association, sexual attraction and incest taboo: A Chinese case. *American Anthropologist, 68,* 883–898.

Wolf, A. P. (1968). Adopt a daughter-in-law, marry a sister: A Chinese solution to the problem of the incest taboo. *American Anthropologist, 70,* 864–874.

Wolf, A. P. (1970). Childhood association and sexual attraction: A further test of the Westermark hypothesis. *American Anthropologist, 72,* 503–515.

Wolf, A. P. (1995). *Sexual attraction and early association: A Chinese brief for Edward Westermark.* Palo Alto, CA: Stanford University Press.

29

Evolutionary Perspectives on Mate Preferences

Diane S. Berry

Children's Medical Center Dallas

After Anna Nicole Smith's untimely death in 2007, many media outlets revisited the headline-grabbing details of her short but highly sensationalized life. In addition to working as a model and actress, the voluptuous blonde captured the title of *Playboy's* "Playmate of the Year" in 1994. In 1996, she again attracted attention by marrying billionaire oil mogul J. Howard Marshall, a man more than 50 years her senior. Marshall expired the next year, setting the stage for a legal battle that reached the U.S. Supreme Court over his fortune. Although the love match doubtless fueled many ribald jokes, few people were actually bewildered as to why Marshall chose to marry a much younger, physically appealing woman such as Anna Nicole rather than someone similar to himself in age and experience. Likewise, few seriously pondered why the former playmate choose the older, wealthy Marshall as a mate rather than a man similar to herself in age and life experience, perhaps one hailing from the small Texas town she called home. To most, the answers to these questions seemed self-evident.

Such "trade-offs" are not unique to current times, and have occurred throughout history. Consider, for example, Henry VIII notorious habit of trading in one wife for another. Typically, each new acquisition was younger and fairer than the previous spouse. For example, Henry married his first wife, Katherine of Aragon, when she was a youthful 20. However, he later divorced her after the much younger Anne Boleyn caught his eye. After Anne became his wife, Henry had her executed, allowing him to marry the even younger Jane Seymour. At age 49, Henry divorced his next wife, Anne of Cleaves, to wed Catherine Howard, a teenager 30 years his junior. Although we only have historical records on which to depend, those descriptions suggest that whenever Henry took a new wife, she was typically more youthful and more attractive than her predecessor. Interestingly, the king's own physical charms were questionable, especially as he aged; moreover, historical reports suggest that these women were indeed more smitten with his power and wealth than with his appearance and youth. It is worth noting that as Henry's power, status, and influence increased across time, his wives generally became younger and more comely (George, 1986).

Without a doubt, the Smith-Marshall match and Henry VIII revolving-door marital style are extreme examples of human mating behavior. However, the general idea that men prize youth and attractiveness in women, whereas women place a higher priority on men's ambition, power, and resources, has received a great deal of empirical support (e.g., Buss, 1989; Buss & Barnes, 1986; Buss, Shackelford, Kirkpatrick, & Larsen, 2001). Moreover, although there are exceptions, more women marry men who are older than themselves than vice versa, and these age preferences are well documented across cultures and time periods (e.g., Buss, 2003; Kenrick & Keefe, 1992).

Sex differences in preferences for these attributes in a romantic partner are one of the most frequently cited predictions of theories of mate selection derived from an evolutionary perspective. In this chapter, we will explore the theoretical underpinnings of this prediction. Specifically, we will first examine some fundamental concepts of evolutionary psychology (EP). Second, we will consider why evolutionary models yield particular predictions about the determinants of romantic attraction, and review research that provides empirical support for these hypotheses. Finally, some important questions raised by an evolutionary perspective on mate selection are considered.

WHAT IS EVOLUTIONARY PSYCHOLOGY?

Evolutionary psychology is a field of science that studies the roots of human psychology and behavior, including the study of mate preferences. More specifically, EP posits that such psychological characteristics evolved in a similar fashion to our physical and anatomical characteristics. A fundamental tenet of EP is Darwin's (1859) concept of natural selection: (a) During our ancestral past, there were differences among members of the human species that influenced whether those members survived and reproduced, and (b) individuals possessing characteristics that facilitated successful survival and reproduction were better represented in future gene pools than those who did not possess them. (Natural selection also applies to nonhuman species, of course, but people are the focus of this chapter.) "Better represented in future gene pools" is simply another way of saying that if Person A has characteristics and attributes that allow him or her to successfully survive and reproduce, and Person B does *not* possess such attributes, Person A is more likely to have children, grandchildren, and, ultimately, great-, great-, great-grandchildren than is Person B. Thus, surviving, successfully reproducing, and leaving a line of successfully reproducing offspring historically translated into being better represented in the future gene pool.

When many people think of evolutionary processes, or natural selection, they tend to focus on survival selection, often called "survival of the fittest." The doctrine of survival selection argues that the genes of individuals who possess traits that provide them with a survival "edge" over those who lack such traits are better represented in later generations. Let's consider a concrete example. Suppose that in ancestral times, members of the human race had the ability to breathe the oxygen-based composition that constitutes earth's air. Assume also that this ability was genetically transmitted from parent to offspring. Next, suppose that through some genetic fluke or mutation, a lucky few human beings developed the heritable ability to intake and successfully process both air and carbon monoxide. What would happen if earth's atmosphere suddenly changed drastically, and our air was replaced with pure carbon monoxide? Whose genes would be inherited by and thus better represented in future generations? Those who could breathe only air would die. However, those who also possessed the ability to breathe carbon monoxide would survive, giving them the opportunity to reproduce and pass that ability on to their offspring. Future generations of the human race would include only descendants of those people who possessed the ability to breath carbon dioxide; in evolutionary terms, the ability to breathe carbon monoxide would become an evolved physical quality, or '"adaptation" (Buss, 2007).

This example involves a hypothetical anatomical characteristic that, via selection, becomes prevalent in a species due to the fact that it helped them survive in their particular environment. Traditionally, this is how people think of the process of evolution. As noted, however, evolutionary psychology goes a step further and posits that *psychological* characteristics—such as preferences, desires, emotions, likes, and dislikes—may also be selected in a similar fashion (i.e., by being passed on to offspring and, ultimately, becoming typical of the population; e.g., Buss, 2007). For example, ancestral individuals who preferred to seek nourishment from pebbles and dirt rather than fruit and meat did not survive. Thus, such dietary preferences would not be carried into future generations, whereas preferences for food and meat of course would, and are with us today.

Note that survival alone does not guarantee that one will contribute to the gene pool. That is, those people who lived to a ripe old age but had no children did not become the ancestors of future generations, whereas those who both survived and reproduced did. Thus, in addition to survival selection, Darwin (1859, 1871) proposed a second important concept, that of sexual selection. In particular, the doctrine of sexual selection states that individuals who possess characteristics that increase their likelihood of successful reproduction (i.e., mating) will be better represented in future gene pools than those who don't. Therefore, sexual selection emphasizes differential reproduction as opposed to survival, and predicts that those heritable traits and characteristics that increase reproductive or mating success (as opposed to survival success) will be "selected" (in evolutionary speak) and also become widespread in the ensuing population. Within EP, sexual selection is posited to be one of the driving forces behind a number of human behaviors and characteristics, including mate preferences (e.g., Buss, 2003, 2007).

Let's consider an illustration of how sexual selection works. In the animal world, for example, peahens prefer peacocks with large, colorful tails to those with less impressive plumages. As females prefer to mate with males who have large, bright tails, those males are most likely to reproduce and pass their genes on to the next generation. Thus, sexual selection favors these males. Note that this differs from survival selection, as the plumage provides no *survival* value, per se. This is a case in which certain individuals are favored genetically due to characteristics that are not directly related to survival, but increase the likelihood of successful reproduction.

Thus, sexual selection can explain why a characteristic not directly related to survival can be selected and become predominant in a species. But why do peahens find these males especially attractive? In fact, large and brilliant plumage in peacocks is an "honest advertisement" of the health and genetic fitness of the male bird. In particular, the more spectacular the plumage, the more of the body's physiological resources must be diverted from other basic functions to produce and maintain it. Therefore, the healthier and more genetically fit the male bird, the more showy the plumage he can produce. Less fit and healthy birds simply can't afford the physiological costs of producing the spectacular tail; in this case, those males' more dull plumage is also an "honest advertisement" of their lesser fitness and poorer health (e.g., Hamilton & Zuk, 1982). Thus, peahens that select mates with impressive plumages produce healthier and more fit offspring than peahens that select males with more drab feathers. In fact, researchers have proposed that male physical characteristics that females find attractive in a variety of species reveal good health and genetic fitness (e.g., Johnson, Thornhill, Ligon, & Zuk, 1993; Moller, 1990).

Do the female birds "know," on some conscious level, that males' showy tails translate into health and fitness, and then consciously figure that into their mating choices? Of course not. What peahens do possess, however, is an evolved mating preference for such plumages. At some point in present-day peahens' ancestral past, females who happened to select such males as mates had greater reproductive success than did their peers. In particular, they produced healthier offspring who, in turn, had a greater likelihood of reproducing successfully. Because of their increased reproductive success, these ancestral peahens were, over time, better represented in the gene pool, and eventually their mating preferences became characteristic of the female of the species.

HUMAN MATE PREFERENCES

What do peahens' preferences and peacocks' tails have to do with Anna Nicole Smith or Henry VIII? According to EP, quite a bit. In fact, evolutionary psychologists have extended the principle of sexual selection to human mate preferences, and argue that sex differences in such preferences are the end result of processes similar to those just described. One of the best-known theoretical accounts of evolved sex differences in preferences for qualities such as age, attractiveness, and resources is sexual strategies theory (SST), originally articulated by Buss and his associates (e.g., Buss, 1998; Buss & Schmitt, 1993).

Buss (2003) suggests that, in general, the ancestral men and women from whom we evolved faced similar "adaptive problems," or challenges to survival and reproduction. As a result, men and women have evolved in a similar fashion within the majority of domains. For example, one adaptive problem involved ingesting sufficient and appropriate nutrients. If an individual successfully solves this challenge, she or he may survive; if not, she or he will die. As noted earlier, people who preferred fruit and meat survived, whereas people who liked to eat pebbles and dirt did not, regardless of their sex. Those men and women passed on their preferences to their offspring. Thus, over time, men and women evolved similar food preferences.

In the reproductive domain, however, the sexes faced somewhat different challenges. In particular, reproductive success (RS), the driving force behind sexual selection, is attained through different avenues for men and women. (Keep in mind that the term "reproductive success" means "success" in the very narrow sense of the extent to which an individual ultimately contributes to the gene pool.) According to SST, these differences ultimately yielded different evolved mate preferences in men and women.

Let's think about the factors that could produce individual differences in people's reproductive success. One influence on RS is the number of children one has. In part, number of offspring influences RS for purely mathematical reasons. For example, if ancestral woman A had 100 children, she would achieve greater reproductive success than woman B, who had 1 child, simply because woman A's children (and their descendants) would account for a greater proportion of the population than would the child and descendants of woman B. Thus, by definition, woman A's genes would be better represented in future generations. A related reason that offspring quantity influences RS is a bit more subtle. In ancestral times, environmental conditions were harsh, and medical interventions virtually nonexistent. As a result, the proportion of infants who survived long enough to mature and reproduce was relatively small. Thus, a man who had 10 offspring would have a better chance of eventually having grandchildren than would a man who had 1 child; the likelihood of having at least some offspring survive and successfully reproduce improved the more offspring one had.

As this illustrates, differences in the quantity of offspring is one factor that influences reproductive success. But note that having numerous offspring is a less effective mechanism of achieving reproductive success for women than for men. In humans, females are more physiologically "invested" in each offspring they produce. That is, although one act of intercourse could be sufficient for a man to produce a child, a woman must, by definition, invest an absolute minimum of nine months to do so. Note that this greater investment limits the quantity of offspring a woman can produce; the maximum number of children she can have during a year's time is one. Thus, producing a high number of offspring is a more difficult road to reproductive success for women than for men. Because the number of offspring an ancestral man could conceivably father during the same time period is much greater, producing high numbers of offspring is indeed a viable road to increasing reproductive success for men.

Based on this concept of differential "parental investment" (Trivers, 1972), evolutionary biologists argue that whichever sex invests most heavily in producing offspring (in humans, females) will be the "choosier" sex when it comes to selecting mates. In a sense, women have fewer opportunities to produce offspring, and selection will favor those who make the most of what opportunities they do have.[1]

If women are comparatively limited in the number of offspring they can produce, what would lead to different levels of female RS? In addition to the quantity of offspring an individual produces, RS is affected by the "quality" or viability of one's offspring. All else being equal, a child who is healthy and genetically fit will be more likely to survive and successfully reproduce than one who is not. Thus, if two women have an equal number of offspring, the RS of the woman with more viable offspring will exceed that of the woman with less viable offspring. It is important to note that viability can be influenced by environmental as well as genetic factors. For example, a child who has access to shelter and nourishment will probably be more likely to survive and reproduce than one who does not.

Thus, one difference between the sexes in the reproductive arena involves how men and women can achieve reproductive success; due to differences in parental investment, women depend more heavily than men on offspring viability;[2] that is, the genes of women who produced more viable offspring, and could secure the resources (e.g., food, shelter, protection, support) to raise them to sexual maturity, were better represented in future generations than women who did not. One implication of this is that women who selected mates who possessed such resources and were willing to invest those resources in her and her offspring ultimately had greater reproductive success than women who chose other mates. Thus, SST posits that preferences for male characteristics that correlate with potential resource accrual evolved over time, and became characteristic of human females.

Men encountered their own adaptive problems in the reproductive arena, although they differed somewhat from those of women (e.g., Buss, 2003). In particular, women's reproductive capacity is directly tied to their age, and there is a relatively narrow "window" of potential fertility in women as compared to men. Whereas men can (theoretically) produce offspring throughout their lifespan, women's ability to produce offspring peaks during the early 20s and then steadily declines during the years leading up to menopause. Thus, identifying fertile mates was more problematic for ancestral men than for women. As a result, SST predicts that men possess evolved preferences for cues to female fertility. These include youth and physical attractiveness, which are directly correlated with fertility. In other words, men who preferred and selected younger, more attractive women as mates experienced greater reproductive success than did those who preferred and chose older, less attractive women. However, because a man's age is less directly tied to his reproductive capacity, "Youth and physical appearance should be less central to a women's mate preferences.…[M]en more than women should value physical attractiveness in mates" (Buss & Schmitt, 1993, p. 209). In fact, because older men were more likely to have accrued wealth and resources, women are actually predicted to have evolved preferences for older as opposed to younger men (Buss, 2003, 2007).

Recall that men and women who had the highest levels of reproductive success became our ancestors; thus, according to EP, we inherited and carry their tendencies to prefer certain qualities in mates. Moreover, because men and women faced different obstacles in achieving reproductive success, men are posited to have essentially inherited somewhat different preferences than women. Thus, SST predicts that present-day women will place a greater value than will men on qualities related to resources and status in a long-term mate, whereas men will place a greater value than women on youth and physical attractiveness in a long-term partner.

Is there any data to support this prediction? The answer seems to be yes. In a seminal study, Buss (1987) analyzed over 10,000 men's and women's self-reported mate preferences in 37 different cultures. Consistent with the predictions of SST, men valued physical attractiveness and youth in a mate more than did women in nearly every culture. On the other hand, women in nearly every culture placed a higher value than did men on mate attributes associated with resource accrual, such as wealth, good financial prospects, industriousness, social status, and ambition. In a reanalysis of preexisting data sets on mate preferences collected from 1939 through 1996, Buss et al. (2001) also found that these sex differences replicate across generations. Other studies of self-reported preferences yield consistent results (e.g., Howard, Blumstein, & Schwartz, 1987; Kenrick, Sadalla, Groth, & Trost, 1990; Regan, Levin, Specher, Christopher, & Cate, 2000; see Okami & Shackelford, 2001, for a review).

Other work replicates these findings using different methodologies. For example, a number of studies have assessed mate preferences by analyzing personal ads and assessing the kinds of qualities men and women claim to seek in a potential partner. In general, women are more likely to indicate a desire for a partner who has appreciable financial resources than are men, and men are more likely to express an interest in a youthful, physically attractive partner than are women. Interestingly, women are also more likely to describe themselves as attractive than are men, and men are more likely to advertise their resources, status, and financial prospects than are women (e.g., Bereczkei, Voros, Gail, & Bernath, 1997; Harrison

[1] This is also the basis of the controversial evolutionary prediction that men evolved preferences for greater sexual variety (i.e., more sexual partners) than women. This is beyond the scope of the present chapter, but interested readers should consult Buss (2003).

[2] Note that it is also true that men's RS would be helped by having healthy, viable offspring. The point is not that men benefit only from offspring quantity or that women benefit only from "quality"; rather, the relative "payoffs," reproductively speaking, are different for each sex.

& Saeed, 1977). Other studies that focused on actual interactions among strangers found that women's attractiveness was a good predictor of the dyad's mutual liking and attraction, whereas men's level of attractiveness was not (Berry & Miller, 2001; Garcia, Stinson, Ickes, Bissonnette, & Briggs, 1991).

Finally, note that the predictions regarding mate preferences made by SST are reflected in real-world mate selections. For example, men who achieve high levels of success tend to marry women who are younger and more physically attractive than the wives of less successful men (e.g., Udry & Eckland, 1984). Cross-culturally, reviews of marital satisfaction reveal that women who marry high-status older men report greater levels of satisfaction, and are less likely to divorce their husbands than are other women (e.g., Bereczkei & Csanaky, 1996; Botwin, Buss, & Shackelford, 1997) As noted previously, numerous studies document that husbands indeed tend to be older than the women they marry (e.g., Buss, Shackelford, & LeBlanc, 2000; Kenrick & Keefe, 1992).

BUT DON'T GIRLS LIKE CUTE GUYS TOO?

In sum, a wealth of evidence supports the predictions made by evolutionary models such as SST regarding mate preferences. However, let's not forget the peahen that preferred the male bird with the "attractive tail" to his competitors. Like the female peahen, human females are not immune to men's physical attractiveness, as many of us might intuitively guess. For example, many studies of the "beautiful is good" effect—the well-established finding that attractive people are preferred over unattractive people in a number of domains—document attractiveness preferences in women, regardless of whether they are making same-sex or opposite-sex judgments (see Langlois et al., 2000, for a review). Even in the studies of self-reported mate preferences just discussed, the data don't indicate that women don't value male attractiveness; they simply don't value attractiveness in a mate to the extent that men do. For example, Buss and Barnes (1986) asked college students to rank-order how important a variety of qualities in a potential mate were to them. Men gave attractiveness a mean ranking of 4.04, and women gave it an average rank of 6.26. The data did support SST's predictions, as men placed more value on attractiveness than did women, whereas women in this study gave higher rankings to male qualities such as status and resources. However, it's clear even from these data that attractiveness plays some role in women's mate preferences.

Because attractiveness is not an indicator of men's ability to reproduce (as male reproduction is not strongly tied to age), what benefit would ancestral woman have gained by choosing physically attractive mates? Sexual strategies theory does not directly address women's attractiveness preferences, but another evolutionary perspective often called "good genes theory" does. These evolutionary models argue that women may have evolved preferences for physically attractive men because attractiveness can signal, or be an "honest indicator" of, genetic fitness (e.g., Thornhill & Gangestad, 1993). If male attractiveness reveals superior fitness, greater reproductive success would have accrued to women who preferred attractive men (via increased offspring "quality" and fitness), thus eventually yielding female preferences for attractiveness in a mate.

Let's consider a specific example of how this might work. Studies have revealed that features such as prominent jaws and chins increase the rated attractiveness of male faces, whereas small, more rounded features are not considered especially attractive (e.g., Berry & McArthur, 1986; Cunningham, Barbee, & Pike, 1990; Keating, 1985). Moreover, these male facial features are secondary sex characteristics that develop due to the influence of testosterone. Testosterone is an immunosuppressant, meaning it decreases the efficiency of the immune system. Thus, these features—like the peacock's tail—are "honest advertisements" of a robust immune system because men with inferior immune systems could not tolerate the physiological stress created by their development. Consequently, women who preferred these traits may have secured greater fitness benefits than their counterparts who preferred other features. Thus, these female preferences may have evolved. Like the peacock's tail, these facial features may also have been selected in men via sexual selection, as they increased mating success (Thornhill & Gangestad, 1993). As with the peacock, this is an example of a physical characteristic evolving due to mating value, *not* survival value; in fact, the development of these features actually taxes the body, which is why they are honest advertisements of fitness (cf. Gangestad & Scheyd, 2005, for a more detailed discussion).

To summarize, the good genes model and SST may initially sound at odds with each other, but they are not. SST simply focuses more directly on the origin of men's preferences for attractiveness, whereas good genes theory focuses on women's interest in attractive men. Therefore, each model focuses on *different* sets of attractiveness-related qualities for which preferences evolved for *different* reasons in men and women. In particular, SST focuses on cues that reveal women's fertility and reproductive status. As the likelihood of a sexual encounter producing offspring is directly linked to a woman's age and, in turn, attractiveness, male preferences for these features increase men's reproductive success. On the other hand, women may find particularly attractive male features that are "honest advertisements" of heritable fitness; this would lead to increased RS via increases in infant fitness and viability. The good genes model focuses on these mechanisms. However, both models may provide important insights into the source of men's and women's mate preferences, specifically in the domain of physical attractiveness.

SOME FOOD FOR THOUGHT

Evolutionary psychology is a broad and expanding field. It's important that readers don't take home the message that all of EP is focused on sex differences in mating preferences, or on the role of physical attractiveness in mate selection, the focus of this chapter. This is only one line of work in evolutionary psychology, albeit an important one that has received a good deal of attention. However, even within the mating arena, evolutionary psychologists study a variety of different topics such as infidelity (e.g., Shackelford, Buss, & Bennett, 2002), jealousy (cf. Buss, 2000), and spousal abuse (e.g., Shackelford, Goetz, Buss, Euler, & Hoier, 2005).

In addition, please don't take away the message that the only topic evolutionary psychologists ever study (or think about) is sex. Because sexual selection is one of the "engines" that drive evolution, mating is, by definition, a major focus of EP. However, evolutionary psychologists study a wide variety of other issues as well. For example, various lines of work in EP are aimed toward developing an understanding of the evolutionary roots of child abuse (e.g., Daly & Wilson, 1998), religion (e.g., Kirkpatrick, 2005), altruism and helping behavior (e.g., Burnstein, Crandall, & Kitayama, 1994; Korchmaros & Kenny, 2001), sibling rivalry (e.g., Sulloway, 2001), anxiety and phobias (e.g., Ohman & Mineka, 2003), homicide (e.g., Buss, 2005; Daly & Wilson, 1988), child development (e.g., Bjorklund & Pellegrini, 2000), and parent/child relations (e.g., Daly, 1990), to name a few.

As wide as the field of evolutionary psychology is, it is also deeply complex, and raises almost as many questions as it answers. The following questions are some that often come up when students encounter a description of evolutionary psychology for the first time. They are good, thoughtful questions. In some cases, EP has a good, thoughtful answer. In other cases, however, these issues are unresolved, and EP has yet to adequately speak to them. Although it is beyond the scope of this chapter to fully address these issues, recommendations for additional readings and sources are provided for interested readers.

Question 1: *Evolutionary psychologists seem to think that 'women are from Venus, men are from Mars.' Is that right?* After reading this chapter, it may seem that way. But evolutionary psychologists who study mating actually stress exactly the *opposite* point. The vast majority of the adaptive problems that ancestral men and women faced were similar: finding nourishment, regulating body temperature, forming coalitions, and seeking protection from predators, to name just a few. Thus, the physiology and psychology of the sexes are remarkably similar, except within the narrow range of mating and reproduction. Perhaps one of the reasons that these few differences seem so salient to us is the emotional weight that mating and reproductive decisions carry. See Buss (2003, 2007) for more on this issue.

Question 2: *EP argues that men look for youth and attractiveness in women because they are 'honest advertisements' of fertility. But today, many young and attractive women use birth control, so preferring them wouldn't increase men's reproductive success. How does EP explain that?* It's important to keep in mind that such preferences theoretically evolved because they solved adaptive problems encountered in our evolutionary *past*, not necessarily during the present. The fact that a particular preference does not make sense today does not mean it did not favor survival and/or reproduction in the lives of our ancestors. Consider the following example. Buss (2007) notes that we have an unhealthy love of fat-laden food. This likely evolved from a very useful preference for such foods in our ancestral past, when these critical resources were difficult to obtain. Now that fatty foods are readily available (i.e., there's a fast-food restaurant on every block), not only does this evolved preference not increase survival, it probably jeopardizes survival. However, evolution is a very slow moving process, and such preferences still exist in the present environment. Similarly, although attractiveness preferences may have no relation to the reproductive success of the current cohort of humans, such desires are still deeply embedded in the human psyche.

Question 3: *When I think of a genetically inherited trait, I think of something like eye color. Does EP really mean to claim that what we like in the opposite sex is inherited by our children? If not, how could these preferences evolve?* In fact, that is exactly the position endorsed by EP. Moreover, the logic of these models would fall apart if such psychological preferences were not somehow genetically specified. For example, Buss (1989) specifically notes that "an adaptation must have genes 'for' that adaptation. Those genes are required for the passage of the adaptation from parents to children; hence, the adaptation can be inherited" (p. 36). (Recall that "adaptation" is another term for a physical or psychological quality that has evolved via selection; thus, mate preferences are considered an adaptation.)

This is a controversial point. However, the idea that attractiveness preferences are hardwired does have some empirical support. For example, despite the fact that many people believe that "beauty is in the eye of the beholder," research reveals that people agree in their judgments of attractiveness, and this agreement is observed across perceivers from very different cultures (cf. Langlois et al, 2000). Moreover, even young infants show a preference for attractive faces (e.g., Langlois, Roggman, & Rieser-Danner, 1990). Nevertheless, many psychologists have questions about this tenet of evolutionary psychology, and would welcome more discussion of evidence for the genetic and heritable basis of likes or dislikes, such as preferences for "men with resources" (see Berry, 2000; Eagly & Wood, 1999, for further discussion).

Question 4: *I don't understand why people's mating decisions are called strategies. Does this mean people really plan how to increase their reproductive success?* No, and the choice of the term "strategies" is (in this author's opinion) an unfortunate one because it has misleading implications. Recall the ancestral peahen that happened to be attracted to the males of her species who displayed the most showy and brilliant tails. She certainly was not calculating reproductive success and making her mating decisions on that basis. She just liked what she liked, and as a fortuitous result, she had more descendants than did her peers. Similarly, consider the ancestral women who happened to favor men with broad jaws. Were they calculating their contribution to the future gene pool when they chose a mate? Of course not. Again, they just followed their hearts (i.e., mate preferences). If that preference happened to increase RS, it would influence the preferences of future women. In short, evolution is not an active process in terms of the decisions of members of a species; people (or peahens) simply make the choices that they make. If those choices happen to increase reproductive success, their preferences evolve. There is no implication that the decisions makers had any "agenda" or plan in mind. They simply acted on their feelings. Evolutionary processes are, in this respect, quite unconscious and passive.

SUMMARY

In sum, evolutionary psychology is a new and fascinating science. Readers are encouraged to not simply dismiss EP because it raises questions. Rather, please explore the issues and questions raised (or others that have occurred to you), perhaps using the references and recommended readings listed. And, by the way, why might it be that Anna Nicole Smith married a much older man when she could (and probably did) easily attract younger men? Why did Henry the Eighth trade one wife for another as he aged? Why did women marry him despite the fact that he was older and less attractive than they were? Perhaps evolutionary psychology brings light to answers to some of these questions. Or, perhaps not. Reader, think; you decide.

REFERENCES AND FURTHER READINGS

Bereczkei, T., & Csanaky, A. (1996). Evolutionary pathway of child development: Lifestyles of adolescents and adults from father absent families. *Human Nature, 7*, 257–280.

Bereczkei, T., Voros, S., Gail, A., & Bernath, L. (1997). Resources, attractiveness, family commitment: Reproductive decisions in human mate choice. *Ethology, 103*, 681–699.

Berry, D. S. (2000). Attractiveness, attraction, and sexual selection: Evolutionary perspectives on the form and function of physical attractiveness. *Advances in Experimental Social Psychology, 32*, 273–342.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial characteristics on social perception. *Psychological Bulletin, 100*, 3–18.

Berry, D. S., & Miller, K. M. (2001). When boy meets girl: Attractiveness and the five-factor model in opposite-sex interactions. *Journal of Research in Personality, 35*, 62–77.

Bjorklund, D. F., & Pellegrini, A. D. (2000). Child development and evolutionary psychology, *Child Development, 71*, 1687–1708.

Botwin, M. D., Buss, D. M., & Shackelford, T. K. (1997). Personality and mate preferences: Five factors in mate selection and marital satisfaction. *Journal of Personality, 65*, 107–136.

Burnstein, E., Crandall, C., & Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology, 67*, 733–789.

Buss, D. M. (1989). Sex differences in human mate selection: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*, 1–49.

Buss, D. M. (1998). Sexual strategies theory: Historical origins and current status. *Journal of Sex Research, 35*, 19–31.

Buss, D. M. (2000). *The dangerous passion: Why jealousy is as necessary as love and sex*. New York: Free Press.

Buss, D. M. (2003). *The evolution of desire: Strategies of human mating* (2nd ed.). New York: Basic Books.

Buss, D. M. (2005). *The murderer next door: Why the mind is designed to kill*. New York: Penguin Press.

Buss, D. M. (2007). *Evolutionary psychology: The new science of the mind* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Buss, D. M., & Barnes, M. F. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50*, 559–570.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*, 204–232.

Buss, D. M., Shackelford, T. K., Kirkpatrick, L. A., & Larsen, R. J. (2001). A half century of mate preferences: The cultural evolution of values. *Journal of Marriage and the Family, 63*, 491–503.

Buss, D. M., Shackelford, T. K., & LeBlanc, G. J. (2000). Number of children desired and preferred spousal age difference: Context-specific mate preference patterns across 37 cultures. *Evolution and Human Behavior, 21*, 323–331.

Cunningham, M. R., Barbee, A. P, & Pike, C. L. (1990) What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology, 59*, 61–72.

Daly, M. (1990). Evolutionary theory and parental motives. In N. A. Krasnegor & R. S. Bridges (Eds.), *Mammalian parenting* (pp. 25–39). New York: Oxford University Press.

Daly, M., & Wilson, M. (1988). *Homicide*. Hawthorne, NY: Aldine.

Daly, M., & Wilson, M. (1998). *The truth about Cinderella: A Darwinian view of parental love*. New Haven, CT: Yale University Press.

Darwin, C. (1859). *On the origin of the species*. London: Murray Press.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: Murray Press.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social rules. *American Psychologist, 54*, 408–423.

Gangestad, S. W., & Cousins, A. J. (2001). Adaptive design, female mate preferences, and shifts across the menstrual cycle. *Annual Review of Sex Research, 12*, 145–185.

Gangestad, S. W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annual Review of Anthropology, 34*, 523–548.

Gangestad, S. W., & Simpson, J. S. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences, 23*, 573–644.

Garcia, S., Stinson, L., Ickes, W., Bissonnette, V., & Briggs, S. (1991). Shyness and attractiveness in mixed sex dyads. *Journal of Personality and Social Psychology, 61*, 35–49.

George, M. (1986). *The autobiography of Henry VIII*. New York: St. Martin's Press.

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science, 218*, 384–387.

Harrison, A. A., & Saeed, L. (1977). Let's make a deal: An analysis of revelations and stipulations in lonely hearts advertisements. *Journal of Personality and Social Psychology, 35*, 257–264.

Howard, J. A., Blumstein, P., & Schwartz, P. (1987). Social or evolutionary theories? Some observations on preferences in human mate selection. *Journal of Personality and Social Psychology, 53*, 194–200.

Johnson, K., Thornhill, R., Ligon, J. D., & Zuk, M. (1993). The direction of mothers' and daughters' preferences and the heritability of male ornaments in the red jungle fowl (*Gallus gallus*). *Behavioral Ecology, 4*, 254–259.

Keating, C. F. (1985). Gender and the physiognomy of dominance and attractiveness. *Social Psychology Quarterly, 48*, 61–70.

Kenrick, D. T., & Keefe, R. C. (1992). Age preferences in mates reflect sex differences in reproductive strategies. *Behavioral and Brain Sciences, 15*, 75–133.

Kenrick, D. T., Sadella, E. K., Groth, G., & Trost, M. R. (1990). Evolution, traits, and the stages of human courtship: Qualifying the parental investment model. *Journal of Personality, 58*, 97–116.

Kirkpatrick, L. A. (2005). Evolutionary psychology: An emerging new foundation for the psychology of religion. In R. F. Paloutzian & C. L. Park (Eds.), *Handbook of the psychology of religion and spirituality* (pp. 101–119). New York: Guilford.

Korchmaros, J. D., & Kenny, D. A. (2001). Emotional closeness as a mediator for the effect of genetic relatedness on altruism. *Psychological Science, 12*, 262–265.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*, 390–423.

Langlois, J. H., Roggman, L. A., & Rieser-Danner, L. A. (1990). Infants' differential social responses to attractive and unattractive faces. *Developmental Psychology, 26*, 153–159.

Moller, A. P. (1990). Fluctuating asymmetry in male sexual ornaments may reliably reveal animal quality. *Animal Behavior, 40*, 1185–1187.

Ohman, A., & Mineka, S. (2003). The malicious serpent: Snakes as a prototypical stimulus for an evolved module of fear. *New Directions in Psychological Science, 12*, 5–9.

Okami, P., & Shackelford, T. K. (2001). Human sex differences in sexual psychology and behavior. *Annual Review of Sex Research, 12*, 186–241.

Regan, P. C., Levin, L., Sprecher, S., Christopher, F. S., & Cate, R. (2000). Partner preferences: What characteristics do men and women desire in their short-term and long-term romantic partners? *Journal of Psychology and Human Sexuality, 12*, 1–21.

Shackelford, T. K., Buss, D. M., & Bennett, K. (2002). Forgiveness or breakup: Sex differences in responses to a partner's infidelity. *Cognition and Emotion, 16*, 299–307.

Shackelford, T. K., Goetz, A. T., Buss, D. M., Euler, H. A., & Hoier, S. (2005). When we hurt the ones we love: Predicting violence against women from men's mate retention. *Personal Relationships, 12*, 447–463.

Sulloway, F. (2001). Birth order, sibling competition, and human behavior. In P. S. Davies & H. R. Holcomb (Eds.), *Conceptual challenges in evolutionary psychology: Innovative research strategies* (pp. 39–83). Dordrecht, Netherlands: Kluwer.

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature, 8*, 237–269.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man* (pp. 136–179). Chicago: University of Chicago Press.

Udry, J. R., & Eckland, B. (1984). Benefits of being attractive: Differential payoffs for men and women. *Psychological Reports, 54*, 47–56.

30

Animal Learning and Behavior

Stephen B. Klein

Mississippi State University

Why do psychologists train rats or monkeys to press a bar for food, or present a buzzer prior to food presentation for cats or dogs, since these situations bear little resemblance to the real world? (In natural settings, rats and monkeys do not have to bar press for food, and cats and dogs do not usually hear a buzzer before they eat.) The answer to this question lies in the belief that there are some general laws of learning. These laws reveal themselves in the study of any behavior, even behaviors not exhibited in natural settings.

Psychologists investigating operant conditioning use the bar press response because many different species acquire it easily. Actually, the unnaturalness of bar pressing is thought to be desirable because the animal comes into the conditioning situation without any past experience that may affect its behavior. The following statement by Skinner (1938) illustrates the belief that the study of any behavior reveals specific laws governing the operant conditioning: "The general topography of operant behavior is not important, because most if not all specific operants are conditioned. I suggest that the dynamic properties of operant behavior may be studied with a single reflex" (pp. 45–46).

Although Skinner studied operant conditioning using the bar press response, the rules Skinner detailed governing the acquisition and extinction of the bar pressing response control the operant conditioning process with many different behaviors and in many species. Thus, using a maze to study the instrumental conditioning process would demonstrate the same rules governing the acquisition or extinction of a bar press response (operant response). Operant conditioning research also demonstrates that different types of reinforcers increase the rate of bar pressing and that the operant conditioning principles identified by Skinner govern behavior in both laboratory and real-world settings. It is not surprising that psychologists have felt confident that training rats and primates to bar press for food reveals the general laws of operant conditioning.

Similarly, psychologists who present a buzzer prior to food assume that any rules they uncover governing the acquisition or extinction of a conditioned salivation response will represent the general laws of classical conditioning. The choice of a buzzer and food is arbitrary: Cats or dogs could be conditioned to salivate as readily to a wide variety of visual, auditory, or tactile stimuli. The following statement by Pavlov (1928) illustrates the view that all stimuli are capable of becoming conditioned stimuli: "Any natural phenomenon chosen at will may be converted into a conditioned stimulus…any visual stimulus, any desired sound, any odor, and the stimulation of any part of the skin" (p. 86).

The specific UCS used also is arbitrary: Any event that can elicit an unconditioned response can become associated with the environmental events that precede it. Thus, Pavlov's buzzer could have as easily been conditioned to elicit fear by pairing it with shock as it was to salivation by being presented prior to food. Pavlov (1928) described the equivalent associability of events in the following statement: "It is obvious that the reflex activity of any effector organ can be chosen for the purpose of investigation, since signaling stimuli can get linked up with any of the inborn reflexes" (p. 17). Pavlov found that many different stimuli can become associated with the UCS of food. Other psychologists documented the conditioning of varied stimuli with a multitude of UCSs. Also, the literature points out that different CSs and UCSs can become associated in both laboratory and natural situations. The idea that any environmental stimulus can become associated with any unconditioned stimulus seemed a reasonable conclusion, based on the research conducted on classical conditioning.

A BEHAVIOR SYSTEMS APPROACH

> All organisms…possess the basic behavioral patterns that enable them to survive in their niches, but learning provides the fine tuning necessary for successful adaptation. (Garcia & Garcia y Robertson, 1985)

The "general laws of learning" view just described assumes that learning is the primary determinant of how an animal acts. According to this approach, learning functions to organize reflexes and random responses so that an animal can effectively interact with the environment. The quote presented above provides a different perspective on the impact of learning on behavior. Rather than assuming that learning organizes behavior, Garcia and Garcia y Robertson (1985) suggested that the organization of behavior already exists within the animal. The function of learning is to enhance already existing organization rather than to create a new organization.

Timberlake's (2001; Timberlake & Lucas, 1989) behavior systems approach suggests that learning modifies preexisting instinctive systems rather than constructing a new behavioral organization. According to Timberlake, an animal possesses a set of instinctive behavior systems such as feeding, mating, social bonding, care of young, and defense. These instinctive behavior systems are independent and serve a specific function or need within the animal.

The predatory subsystem of the feeding system in rats illustrates this view. The predatory sequence begins with the rat in a general search mode. The search mode causes the rat to show enhanced general searching, which leads to greater locomotion and increased sensitivity to spatial and social stimuli that are likely to bring the rat closer to food. The increased locomotion and greater sensitivity to the environment lead the rat to notice a small moving object in the distance as its prey. Once the prey is close, the rat shifts from a general search mode to a focal search mode. The focal search mode causes the rat to engage in a set of perceptual-motor modules related to capturing and subduing its prey. Once the prey is captured, the rat shifts to a handle/consume mode, which elicits the biting, manipulating, chewing, and swallowing involved in the consumption of the prey. This complex instinctive predatory behavior system allows the animal to find and consume the nutrients it needs to survive, with the search modes motivating the perceptual-motor modules that allow the rat to locate, approach, and consume its prey.

In Timberlake's behavior systems approach, learning changes the integration, tuning, instigation, or linkages within a particular behavior system. For example, a new environmental stimulus could become able to release an instinctive motor-response module as a result of a Pavlovian conditioning experience. Learning can also alter the intensity of a simple motor response due to repetition, or improve the efficiency of a complex behavior pattern as a result of the contingent delivery of a reinforcer.

Cues that signal the receipt of reinforcers can be conditioned to activate a mode. For example, cues associated with the prey can be conditioned to the search modes that bring the rat into contact with its prey. However, the conditioning of modes is different from the conditioning of perceptual-motor modules (Timberlake, 2001). Specific motor responses are conditioned to specific stimuli as a result of the conditioning of perceptual-motor modules. By contrast, the conditioning of a specific mode produces a general motivational state that sensitizes all of the perceptual-motor modules in that mode. For example, activation of the general search mode of the predatory subsystem sensitizes the rat to engage in the travel, socialize, investigate, chase, and lie-in-wait perceptual-motor modules.

Timberlake's behavior systems approach assumes that different stimuli can be conditioned to different modes. Distal cues are relevant to a general search mode—the rat must locate distant prey. By contrast, proximal cues are relevant to a focal search mode—the rat must capture and subdue its prey. Silva, Timberlake, and Gont (1998) evaluated the view that distant cues are associated with a general search mode, whereas proximal cues are associated with a focal search mode. In their study, there were two levers on each side of a food tray—one lever was "far" from the food tray, the other "near" to it. Each lever was presented in succession for four seconds, followed by food. In the F-N condition, the far (F) lever was presented first, followed by the near lever; in the N-F condition, the near (N) lever was presented first, followed by the far lever. Silva et al. found that animals in the F-N condition first attended to the far lever, then transferred their attention to the near lever, and finally nosed the food tray just prior to the presentation of food. In Timberlake's view, the F-N condition resembles the predatory subsystem: The general search mode activated attention to the distal far lever, then the focal search mode activated attention to the proximal near lever. But what about the N-F condition? Silva et al. found that the N-F animals attended first to the near lever and then to the food tray, but showed no response to the far lever. The second (far) lever led the rat away from food; because activation of the focal search mode by the first near lever focused attention toward and not away from food, no conditioning occurred to the distal far lever. In fact, rats in the N-F condition spent more time nosing the first near lever than did rats in the F-N condition.

One of the functional features of the behavior system approach is that variations in learning occur between species (Timberlake, 2001). In Timberlake's view, different species of animals learn a particular behavior at different rates. Different species also learn different ways of responding to a particular situation. Considerable variation also occurs within a species. Timberlake suggested that some behaviors are learned more rapidly than others within a given species. Further, there may be different rates of learning of a particular behavior between different members of a species.

What causes these variations between and within animal species? Timberlake (2001) proposed that the variations are due to either predispositions or constraints on what an animal or a person can learn. A predisposition refers to instances in which an animal learns more rapidly or in a different form than expected. Predispositions occur when environmental circumstance easily modifies the instinctive behavior system that the animal brings into the situation. Variations in learning also can reflect the impact of a constraint on learning; a constraint occurs when an animal learns less rapidly or less completely than expected. Constraints on learning occur when environmental circumstance is not suited to the animal's instinctive behavior system. In the next five sections, we will examine examples of predispositions and constraints on learning.

ANIMAL MISBEHAVIOR

Keller Breland and Marian Breland (1961) wanted to see if operant procedures could be used to teach exotic behaviors to animals. They trained 38 species, including reindeer, cockatoos, raccoons, porpoises, and whales, at Animal Behavior Enterprises in Hot Springs, Arkansas. In fact, they trained over 6,000 animals to emit a wide range of behaviors, including teaching hens to play a five-note tune on a piano and perform a "tap dance," pigs to turn on a radio and eat breakfast at a table, chicks to run up an inclined platform and slide off, a calf to answer questions in a quiz show by lighting either a "yes" or "no" sign, and two turkeys to play hockey. These exotic behaviors have been on display at many municipal zoos and museums of natural history, in department store displays, at fair and trade convention exhibits, tourist attractions, and on television. These demonstrations have not only provided entertainment for millions of people but have also documented the power and generality of the operant conditioning procedures Skinner described.

Although K. Breland and M. Breland (1961, 1966) were able to condition a wide variety of exotic behaviors using operant conditioning, they noted that some operant responses, although initially performed effectively, deteriorated with continued training despite repeated food reinforcements. According to K. Breland and M. Breland, the elicitation of instinctive food-foraging and food-handling behaviors by the presentation of food caused the decline in the effectiveness of an operant response reinforced by food. These instinctive behaviors, strengthened by food reinforcement, eventually dominated the operant behavior. They called the deterioration of an operant behavior with continued reinforcement *instinctive drift*, and the instinctive behavior that prevented the continued effectiveness of the operant response *animal misbehavior*. One example of animal misbehavior is described next.

K. Breland and M. Breland (1961, 1966) tried to get pigs to pick up a large wooden coin and deposit it in a piggy bank several feet away. Depositing four or five coins was needed to receive one food reinforcement. According to the Brelands, "Pigs condition very rapidly, they have no trouble taking ratios, they have ravenous appetites (naturally), and in many ways are the most trainable animals we have worked with." However, each pig exhibited an interesting pattern of behavior following conditioning. At first, the pigs picked up a coin, carried it rapidly to the bank, deposited it, and readily returned for another coin. However, over a period of weeks, the pigs' operant behavior became slower and slower. Each pig still rapidly approached the coin, but rather than carry it immediately over to the bank, the pigs "would repeatedly drop it, root it, drop it again, root it along the way, pick it up, toss it up in the air, drop it, root it some more, and so on."

Why did the pigs' operant behavior deteriorate? According to K. Breland and M. Breland (1961, 1966), the pigs merely exhibited the instinctive behaviors associated with eating. The presentation of food not only reinforces the operant response, but it also elicits instinctive food-related behaviors. The reinforcement of these instinctive food-gathering and food-handling behaviors strengthens the instinctive behaviors, which results in the deterioration of the pigs' operant responses (depositing the coin in the bank). The more dominant the instinctive food-related behaviors become, the longer it takes for the operant response to occur. The slow deterioration of the operant depositing response provides support for their instinctive drift view of animal misbehavior.

K. Breland and M. Breland (1961, 1966) have observed many other instances of animal misbehavior. They found hamsters that stopped responding in a glass case, porpoises and whales that swallowed balls or inner tubes instead of playing with them to receive reinforcement, cats that refused to leave the area around the food dispenser, and rabbits that refused to approach their feeder. They also reported extreme difficulty in conditioning many bird species to vocalize to obtain food reinforcement. In each case of animal misbehavior, they suggested that the instinctive food-seeking behavior prevented the continued high performance level of an operant response required to receive reinforcement. These findings suggest that the effectiveness of food reinforcement to establish an operant behavior is limited.

Boakes, Poli, Lockwood, and Goodall (1978) established a procedure for producing animal misbehavior in a laboratory. These researchers trained rats to press a flap to obtain a ball bearing and to deposit it in a chute to obtain food reinforcement. They reported that although all the rats initially released the ball bearing readily, the majority of the animals became reluctant to let go of the ball bearing after several training sessions. These rats repeatedly mouthed, pawed, and retrieved the ball bearing before finally depositing it in the chute.

K. Breland and M. Breland (1961, 1966) suggested that the elicitation and strengthening of instinctive food-related behaviors during operant conditioning is responsible for animal misbehavior. Boakes et al. (1978) proposed another explanation. In their view, animal misbehavior is produced by Pavlovian conditioning rather than by operant conditioning. The association of environmental events with food during conditioning causes these environmental events to elicit species-typical foraging and food-handling behaviors; these behaviors then compete with the operant behavior. Consider the misbehavior of the pig detailed earlier. According to K. Breland and M. Breland, the pigs rooted the tokens because the reinforcement presented during operant conditioning produced and strengthened the instinctive food-related behavior; in contrast, Boakes et al. suggested that the association of the token with food caused the token to elicit the rooting behavior.

Timberlake, Wahl, and King (1982) conducted a series of studies to evaluate the validity of each view of animal misbehavior. The results of the experiments Timberlake et al. conducted show that both operant and Pavlovian conditioning contribute to producing animal misbehavior. In their appetitive structure view, pairing food with the natural cues controlling food-gathering activities produces species-typical foraging and food-handling behaviors. The instinctive food-gathering behaviors must be reinforced if misbehavior is to dominate the operant behavior. Animal misbehavior does not occur in most operant conditioning situations because (a) the cues present during conditioning do not resemble the natural cues eliciting instinctive foraging and food-handling behaviors, and (b) these instinctive behaviors are not reinforced.

Timberlake et al. (1982) used the ball bearing procedure that Boakes et al. (1978) had developed to validate their appetitive structure view of animal misbehavior. Recall that rats in this situation repeatedly mouth, paw, and retrieve the ball bearing before releasing it down the chute to obtain food reinforcement. In Experiment 1, they assessed the contribution of Pavlovian conditioning to animal misbehavior by pairing the ball bearing with food in experimental subjects. Experimental treatment animals received food after the ball bearing had rolled out of the chamber. This study also used two control conditions to evaluate the importance of pairing the ball bearing and food: Animals in one control condition were given random pairings of the ball bearing and food (random group); subjects in the second control condition received only the ball bearing (CS-only group). They reported that experimental group animals exhibited a significant amount of misbehavior toward the ball bearing: they touched it, carried it about the cage, placed it in their mouths, and bit it while holding it in their forepaws. By contrast, infrequent misbehavior occurred in animals in the two control groups. These observations indicate that the ball bearing and food must be presented together for a high level of misbehavior to occur.

The pairing of the ball bearing with food is necessary but not sufficient for the development of misbehavior. Timberlake et al. (1982) asserted that the misbehavior must be reinforced by food presentation for misbehavior to dominate operant responding. Experiments 3 and 4 evaluated the importance of operant conditioning to the establishment of animal misbehavior. In Experiment 3, contact with the ball bearing caused food to be omitted. If reinforcement of contact with the ball bearing is necessary for the dominance of animal misbehavior, the contingency that contact with the ball bearing would prevent reinforcement should lead to an absence of animal misbehavior. The results of Experiment 3 show that if contact with the ball bearing prevents reinforcement, then the animals did indeed exhibit no contact with the ball bearing. In Experiment 4, the contact with the ball bearing was reinforced. If the animal did not touch the ball bearing, it received no food on that trial. Timberlake et al. reported that reinforcement of contact with the ball bearing produced a rapid increase in the level of animal misbehavior. These studies suggest that for misbehavior to develop, stimuli (e.g., ball bearings) resembling the natural cues controlling food-gathering activities must be consistently paired with food (Pavlovian conditioning), and the presentation of food must reinforce the occurrence of species-typical foraging and food-handling behaviors elicited by the natural cues (operant conditioning).

SCHEDULE-INDUCED BEHAVIOR

B. F. Skinner (1948) described an interesting pattern of behavior that pigeons exhibited when reinforced for key pecking on a fixed-interval schedule. When food reinforcement was delivered to the pigeons on a fixed-interval, 15-second schedule, they developed a "ritualistic" stereotyped pattern of behavior during the interval. The pattern of behavior differed from bird to bird—some walked in circles between food presentations; others scratched the floor; still others moved their heads back and forth. Once a particular pattern of behavior emerged, the pigeons repeatedly exhibited it, with the frequency of the behavior increasing as the birds received more reinforcement. Skinner referred to the behaviors of his pigeons on the interval schedule as examples of superstitious behavior.

Why do animals exhibit superstitious behavior? One reasonable explanation suggests that animals have associated the superstitious behavior with reinforcement, and this association causes the animals to exhibit high levels of the superstitious behavior. Staddon and Simmelhag's (1971) analysis of superstitious behavior indicated that it is not an example of the operant behavior. They identified two types of behavior produced when reinforcement (for example, food) is programmed to occur on a regular basis: *terminal behavior* and *interim behavior*. Terminal behavior occurs during the last few seconds of the interval between reinforcer presentations, and it is reinforcer oriented. Pigeons peck on or near the food hopper that delivered food; this is an example of terminal behavior. Interim behavior, in contrast, is not reinforcer oriented. Although contiguity influences the development of terminal behavior, interim behavior does not occur contiguously with reinforcement. Terminal behavior falls between interim behavior and reinforcement but does not interfere with the exhibition of interim behavior.

Staddon and Simmelhag (1971) suggested that terminal behavior occurs in stimulus situations that are highly predictive of the occurrence of reinforcement—that is, terminal behavior is typically emitted just prior to reinforcement on a fixed-interval schedule. By contrast, interim behavior occurs during stimulus conditions that have a low probability of the occurrence of reinforcement—that is, interim behavior is observed most frequently in the period following reinforcement.

The strange, superstitious behavior that Skinner initially described is only one example of interim behavior. Animals exhibit a wide variety of other behaviors (e.g., drinking, running, grooming, nest building, aggression) when reinforcement occurs regularly. When fixed-interval schedules of reinforcement elicit high levels of interim behavior, we refer to it as *schedule-induced behavior*.

Schedule-Induced Polydipsia

The most extensively studied form of schedule-induced behavior is the excessive intake of water (polydipsia) when animals are reinforced with food on a fixed-interval schedule. John Falk (1961) conducted a study to observe schedule-induced polydipsia. He deprived rats of food until their body weight was approximately 70 to 80 percent of their initial weight and then trained them to bar press for food reinforcement. When water was available in an operant chamber, Falk found that the rats consumed excessive amounts of water. Even though the rats were not water deprived, they drank large amounts of water; in fact, under certain conditions, animals provided food reinforcement on an interval schedule will consume as much as one half their weight in water in a few hours. Not water deprivation, heat stress, or a similar amount of food in one meal produces this level of excessive drinking. Apparently, some important aspect of providing food on an interval schedule can elicit excessive drinking.

Is schedule-induced polydipsia an example of interim behavior? Recall Staddon and Simmelhag's (1971) definition—interim behavior occurs in stimulus situations that have a low probability of reinforcement occurrence. Schedule-induced drinking does fit their definition: Animals reinforced on an interval schedule typically drink during the period following food consumption. In contrast, drinking usually does not occur in the period that precedes the availability of food reinforcement.

Schedule-induced polydipsia has been consistently observed in rats given food on an interval schedule (Wetherington, 1982). A variety of different interval schedules of reinforcement have been found to produce polydipsia. Falk (1961) observed polydipsia in rats on a fixed-interval schedule, and Jacquet (1972) observed polydipsia on a variety of compound schedules of reinforcement. Schedule-induced polydipsia also is found in species other than rats. Shanab and Peterson (1969) reported schedule-induced polydipsia in pigeons, and Schuster and Woods (1966) observed it in primates.

Other Schedule-Induced Behaviors

Several other instinctive behaviors are observed in animals on interval schedules. A number of psychologists (King, 1974; Staddon & Ayres, 1975) have reported that interval schedules of reinforcement produce high levels of wheel running. Schedule-induced wheel running was observed using both food and water reinforcement. Levitsky and G. Collier (1968) found that the highest rate of wheel running occurs in the time immediately following reinforcement and then decreases as the time for the next reinforcement nears, while Staddon and Ayres reported that as the interreinforcement interval increases, the intensity of wheel running initially increases and then declines.

Animals receiving reinforcement on an interval schedule will attack an appropriate target of aggressive behavior. Cohen and Looney (1973) reported that pigeons will attack another bird or a stuffed model of a bird present during key pecking for reinforcement on an interval schedule. Similar schedule-induced aggression appears in squirrel monkeys (Hutchinson, Azrin, & Hunt, 1968) and rats (Knutson & Kleinknecht, 1970). Knutson and Kleinknecht reported that the greatest intensity of aggressive behavior occurred in the immediate postreinforcement period.

FLAVOR-AVERSION LEARNING

Contiguity plays a critical role in the acquisition of a conditioned response: Little conditioning occurs if the CS precedes the UCS by several seconds or minutes. However, animals will develop aversions to taste cues even when the taste stimulus preceded illness by several hours (see also Chapter 35). This indicates that, unlike other conditioned responses, the flavor aversion does not depend on contiguity. The association of a flavor with illness is often referred to as long-delay learning; this term suggests a difference between flavor-aversion learning and other examples of classical conditioning.

Some stimuli are more likely than others to become associated with a particular UCS. Garcia and Koelling's (1966) classic study shows that a taste is more salient when preceding illness than when preceding shock, whereas a light or tone is more salient when preceding shock than when preceding illness. In Garcia and Koelling's study, rats were exposed to either a saccharin taste cue or a light-and-tone compound stimulus. Following exposure to one of these cues, animals received either an electric shock or irradiation-induced illness. Animals exhibited an aversion to saccharin when it was paired with illness but not when it was paired with shock. In addition, they developed a fear of the light-and-tone stimulus when it was paired with shock but not when paired with illness.

On the basis of the Garcia and Koelling study, Seligman (1970) proposed that rats have an evolutionary preparedness to associate tastes with illness. Further support for this view is the observation that adult rats acquire an intense aversion to a flavor after a single taste-illness pairing. Young animals also acquire a strong aversion after one pairing (Klein, Domato, Hallstead, Stephens, & Mikulka, 1975). Apparently, taste cues are very salient in terms of their associability with illness.

Seligman also suggested that rats are contraprepared to become afraid of a light or tone paired with illness. However, other research (Klein, Freda, & Mikulka, 1985) indicates that rats can associate an environmental cue with illness. Klein et al. found that rats avoided a distinctive black compartment previously paired with an illness-inducing apormorphine or lithium chloride injection, whereas Revusky and Parker (1976) observed that rats did not eat out of a container that had been paired with illness induced by lithium chloride. Although animals can acquire environmental aversions, more trials and careful training procedures are necessary to establish an environmental aversion than for a flavor aversion (Riccio & Haroutunian, 1977).

Although rats form flavor aversions more readily than environmental aversions, other species do not show this pattern of stimulus salience. Unlike rats, birds acquire visual aversions more rapidly than taste aversions. Wilcoxin, Dragoin, and Kral (1971) induced illness in quail that had consumed sour blue water. They reported that an aversion formed to the blue color, but not to the sour taste. In the same vein, Capretta (1961) found greater salience of visual cues than taste cues in chickens.

Why are visual cues more salient than taste stimuli in birds? According to Garcia, Hankins, and Rusiniak (1974), this salience hierarchy is adaptive. Since birds' seeds are covered by a hard, flavorless shell, they must use visual cues to assess whether food is poisoned; thus, visual cues enable birds to avoid consuming poisonous seeds. Although this view seems reasonable, it does not appear to be completely accurate. Braveman (1974, 1975) suggests that the feeding time characteristic of a particular species determines the relative salience of stimuli becoming associated with illness. Rats, which are nocturnal animals, locate their food at night and therefore rely less on visual information than on gustatory information to identify poisoned food. By contrast, birds search for their food during the day, and visual information plays an important role in controlling their food intake. Braveman evaluated this view by examining the salience hierarchies of guinea pigs, which, like birds, seek their food during the day. He found visual cues to be more salient than taste stimuli for guinea pigs.

IMPRINTING

Infant Love

You have undoubtedly seen young ducks swimming behind their mother in a lake. What process is responsible for the young birds' attachment to their mother? Lorenz (1952/1957) investigated this social attachment process, calling it *imprinting*. Lorenz found that a newly hatched bird approaches, follows, and forms a social attachment to the first moving object it encounters. Although typically the first object that the young bird sees is its mother, birds imprint to many different and sometimes peculiar objects. In a classic demonstration of imprinting, newly hatched goslings imprinted to Lorenz and thereafter followed him everywhere. Birds have imprinted to colored boxes and other inanimate objects as well as to animals of different species. After imprinting, the young animal prefers the imprinted object to its real mother; this shows the strength of imprinting.

Although animals have imprinted to a wide variety of objects, certain characteristics of the object affect the likelihood of imprinting. P. H. Klopfer (1971) found that ducklings imprinted more readily to a moving object than to a stationary object. Also, ducks are more likely to imprint to an object that (a) makes "lifelike" rather than "gliding" movements (Fabricius, 1951); (b) vocalizes rather than remains silent (N. E. Collias & E. C. Collias, 1956); (c) emits short rhythmic sounds rather than long high-pitched sounds (Weidman, 1956); and (d) measures about 10 cm in diameter (Schulman, Hale, & Graves, 1970).

Harry Harlow (1971) observed that baby primates readily became attached to a soft terry cloth surrogate mother but developed no attachment to a wire mother. Harlow and Suomi (1970) found that infant monkeys preferred a terry cloth mother to a rayon, vinyl, or sandpaper surrogate; liked clinging to a rocking mother rather than to a stationary mother; and chose a warm (temperature) mother over a cold one. Mary Ainsworth and her associates (Ainsworth, 1982; Blehar, Lieberman, & Ainsworth, 1977) reported that human infants also need a warm, responsive mother for social attachment; they found a strong attachment to mothers who were responsive and sensitive to their children's needs. By contrast, infants showed little attachment to anxious or indifferent mothers.

Age plays an important role in the imprinting process. Not only does imprinting occur readily during certain sensitive periods, but also imprinting is less likely to occur following this sensitive period. Jaynes's (1956) study exposed newly hatched New Hampshire chicks to cardboard cubes at different times. He reported that five-sixths of the chicks imprinted within 1 to 6 hours after hatching. However, only five-sevenths of the chicks met the criterion for imprinting when exposed to the cardboard cube 6 to 12 hours after hatching. The percentage declined to three-fifths at 24 to 30 hours, two-fifths at 30 to 36 hours, and only one-fifth at 48 to 54 hours. The sensitive period for social attachment differs between species; in sheep and goats, it is 2 to 3 hours after birth (P. H. Klopfer, Adams, & M. S. Klopfer, 1964); in primates, 3 to 6 months; and in humans, 6 to 12 months (Harlow, 1971).

However, the sensitive period merely reflects a lesser degree of difficulty in forming an attachment; when sufficient experience is given, imprinting will occur after the sensitive period has lapsed. Brown (1975) trained ducklings ranging in age from 20 to 120 hours to follow an object to an equivalent degree. He found that although the older the duck, the longer the time required for imprinting to occur, all the ducklings with sufficient training showed an equal degree of attachment to the imprinted object.

Imprinting differs from other forms of associative learning (Davey, 1989). The animal's response to the imprinting object is less susceptible to change than an animal's reaction to events acquired through conventional associative learning is. Conditioned stimuli that elicit saliva quickly extinguish when food is discontinued. Similarly, the absence of shock produces a rapid extinction of fear. By contrast, the elimination of reinforcement does not typically lead to a loss of reaction to an imprinting object. Hess (1962, 1964) observed that once a three- to four-day-old chick developed a food preference to a less preferred object, this preference remained, despite the subsequent lack of food reinforcement when the chick pecked at this object.

Although punishment quickly alters an animal's response to a conditioned stimulus, animals seem insensitive to punishment from an imprinting object. Kovach and Hess (1963) found that chicks approached the imprinting object despite its administration of electric shock. Harlow's (1971) research shows how powerful the social attachment of the infant primate is to its surrogate mother. Harlow constructed four abusive "monster mothers." One rocked violently from time to time; a second projected an air blast in the infant's face. Primate infants clung to these mothers even as they were abused. The other two monster mothers were even more abusive: One tossed the infant off her, and the other shot brass spikes as the infant approached. Although the infants were unable to cling to these mothers continuously, they resumed clinging as soon as possible when the abuse stopped. His observations are consistent with observations of many abused children, who typically desire to return to their abusive parent.

Sexual Preference

Lorenz (1957) reported an interesting behavior in one of his male jackdaws. The bird attempted courtship feeding with him: It finely minced worms, mixed them with saliva, and attempted to place the worms in Lorenz's mouth. When Lorenz did not open his mouth, he got an earful of worm pulp. Lorenz suggested that the male jackdaw had sexually imprinted to him.

The sexual preference of many birds is established during a sensitive period (Eibl-Eibesfeldt, 1970; Lorenz, 1957). Also, the birds' sexual preference does not have to be for their own species; that is, a sexual preference can be established to another species if exposure occurs during the sensitive period. Because sexual preference develops in immature birds when copulation is impossible, the establishment of the birds' sexual preference does not depend upon sexual reinforcement. Further, the imprinted bird's sexual preference is not modified even after sexual experience with another bird species. Perhaps this type of sexual imprinting is a cause of the development and persistence of human sexual preferences.

Food Preference

Hess (1962, 1964) suggested that an animal's experience with food during a sensitive period of development results in the establishment of a food preference. This preference can develop to a typically nonpreferred food and, once established, is permanent. Chicks innately prefer to peck at a white circle on a blue background rather than a white triangle on a green background. Hess gave different groups of chicks of various ages experience with the less-preferred stimulus. The chicks developed a strong preference for the white-triangle-green-background stimulus if they experienced it during days three to four after the chicks had hatched. This preference did not develop if the experience occurred on days one, two, seven, or nine after hatching. These observations indicate that the sensitive period for the establishment of food preference in chicks is three to four days following hatching. Hess suggested that this time period for the establishment of a food preference is critical because three-day-old chicks no longer use the yolk sac for nutrients and can peck with maximum accuracy.

Humans differ considerably in their food preferences (Rozin & Zellner, 1985). These preferences may to some degree reflect experience with a specific food during the sensitive period of development. People typically prefer familiar foods, which suggests an imprinting influence in food preference. Food aversions in humans may also be sensitive to a developmental stage; Garb and Stunkard's (1974) observation that people are most apt to develop food aversions between the ages of 6 and 12 years provides additional evidence that imprinting affects the establishment of food preferences and aversions.

AVOIDANCE OF AVERSIVE EVENTS

Species-Specific Defense Reactions

Bolles (1970, 1978) suggested that animals have species-specific defense reactions (SSDR) that allow them to avoid dangerous events. According to Bolles, animals have little opportunity to learn to avoid danger: They either possess an instinctive means of keeping out of trouble or they perish. For example, a deer does not have time to learn to avoid its predator. Unless the deer possesses an instinct for avoiding predators, it will probably wind up as a predator's meal.

The instinctive responses that enable animals to avoid aversive events differ. An animal's evolutionary history determines which behaviors will become SSDRs: Responses that enable animals to avoid aversive events remain in their genetic programming, whereas nonadaptive responses will not be passed on to future generations. Bolles (1970, 1978) proposed that animals experiencing danger narrow their response repertoire to those behaviors that they expect will eliminate the danger. Because evolution has proved the species-specific defense reactions to be effective and other behaviors likely to produce failure, behaviors other than the species-specific defense reactions probably would be nonadaptive. Thus, animals limit their reactions to SSDRs as they attempt to avoid danger.

Rats employ three different species-specific defense reactions: running, freezing, and fighting. They attempt to run from a distant danger; a close danger motivates freezing. When these two responses fail, rats use aggressive behavior to avoid aversive events. Other animals employ different instinctive responses to avoid danger: The mouse, as Bolles (1970, 1978) suggested in a quote from Robert Burns's "To a Mouse," is "a wee timorous beastie" when experiencing danger because this is the only way this small and relatively defenseless animal can avoid danger. By contrast, the bird just flies away.

Bolles and A. C. Collier (1976) demonstrated that the cues that predict danger not only motivate defensive behavior but also determine which response rats will exhibit when they expect danger. Rats received shock in a square or a rectangular box. After they experienced the shock, the rats either remained in the dangerous environment or were placed in another box where no shocks were given. Defensive behavior occurred only when the rats remained in the compartment where they had previously been shocked. Bolles and A. C. Collier also found that a dangerous square compartment produced a freezing response, while a dangerous rectangular box caused the rats to run. These results suggest that the particular SSDR produced depends on the nature of the dangerous environment.

Animals easily learn to avoid an aversive event when they can use an SSDR. Rats readily learn to run to avoid being shocked. Similarly, pigeons easily learn to avoid shock by flying from perch to perch. By contrast, animals have difficulty learning to avoid an aversive event when they must emit a behavior other than an SSDR to avoid the aversive event. D'Amato and Schiff (1964) trained rats to bar press to avoid electric shock. They reported that over half of their rats, even after having participated in more than 7,000 trials over a four-month period, failed to learn the avoidance response.

Bolles (1969) provides additional evidence of the importance of instinct in avoidance learning. He reported that rats quickly learned to run in an activity wheel to avoid electric shock, but found no evidence of learning when the rats were required to stand on their hind legs to avoid shock. Although rats stood on their hind legs in an attempt to escape from the compartment where they were receiving shock, these rats did not learn the same behavior to avoid shock. Bolles suggested that a rat's natural response in a small compartment was to freeze, and this innate SSDR prevented them from learning a nonspecies-specific defensive reaction as the avoidance behavior.

Predispositions and Avoidance Learning

Bolles (1978) suggested that predispositions are responsible for the development of avoidance learning. In his view, aversive events elicit instinctive species-specific defensive responses. The environment present during aversive events becomes able to produce these instinctive defensive reactions as a conditioned response. Whereas instinctive CRs to cues associated with reinforcement elicit approach and contact behavior that enables an animal to obtain reinforcement, stimuli associated with aversive events produce instinctive defensive responses that allow the animal to avoid aversive events.

Bolles (1978) suggested that Pavlovian conditioning rather than operant conditioning influences whether or not avoidance learning occurs. The association of environmental stimuli with aversive events rather than reinforcement causes the development of avoidance behavior. Bolles and Riley's (1973) study shows that reinforcement is not responsible for the rapid acquisition of avoidance behavior. In their study, some animals could avoid being shocked by freezing. They reported that after only a few minutes of training, their animals were freezing most of the time. Two additional groups were included in their study: One group was punished for freezing and could avoid shock by not freezing; the other group was shocked regardless of their behavior. Bolles and Riley observed that the rats punished for freezing still froze much of the time. Furthermore, rats punished for freezing froze as much as the rats that were shocked regardless of their behavior. Bolles suggested that when an animal is in a small confined area and anticipates an aversive event, it freezes. The animals punished for freezing would still have frozen all the time had frequent shocks not disrupted their freezing. Yet, as soon as the shock ended, the anticipation elicited the instinctive freezing response. Thus, with the exception of shock-induced disruption of freezing in animals punished for freezing, animals either reinforced or punished for freezing showed equivalent levels of freezing behavior. These results suggested that the contingency between the freezing response and the aversive event did not affect these animals' behavior.

REFERENCES AND FURTHER READINGS

Ainsworth, M. D. S. (1982). Attachment: Retrospect and prospect. In C. M. Parkes & J. Sevenson-Hinde (Eds.), *The place of attachment in human behavior.* New York: Basic Books.

Blehar, M. C., Lieberman, A. F., & Ainsworth, M. D. S. (1977). Early face-to-face interaction and its relation to later infant-mother attachment. *Child Development, 48,* 182–194.

Boakes, R. A., Poli, M., Lockwood, M. J., & Goodall, G. (1978). A study of misbehavior: Token reinforcement in the rat. *Journal of the Experimental Analysis of Behavior, 29,* 115–134.

Bolles, R. C. (1970). Species-specific defense reactions and avoidance learning. *Psychological Review, 77,* 32–48.

Bolles, R. C. (1978). The role of stimulus learning in defensive behavior. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 89–108). Hillsdale, NJ: Erlbaum.

Bolles, R. C., & Collier, A. C. (1976). The effect of predictive cues on freezing in rats. *Animal Learning and Behavior, 4,* 6–8.

Bolles, R. C., & Riley, A. (1973). Freezing as an avoidance response: Another look at the operant-respondent distinction. *Learning and Motivation, 4,* 268–275.

Braveman, N. S. (1974). Poison-based avoidance learning with flavored or colored water in guinea pigs. *Learning and Motivation, 5,* 182–194.

Braveman, N. S. (1975). Formation of taste aversions in rats following prior exposure to sickness. *Learning and Motivation, 6,* 512–534.

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist, 61,* 681–684.

Breland, K., & Breland, M. (1966). *Animal behavior.* New York: Macmillan.

Brown, J. L. (1975). *The evolution of behavior.* New York: Norton.

Capretta, P. J. (1961). An experimental modification of food preference in chickens. *Journal of Comparative and Physiological Psychology, 54,* 238–242.

Cohen, P. S., & Looney, T. A. (1973). Schedule-induced mirror responding in the pigeon. *Journal of the Experimental Analysis of Behavior, 19,* 395–408.

Collias, N. E., & Collias, E. C. (1956). Some mechanisms of family integration in ducks. *Auk, 73,* 378–400.

D'Amato, M. R., & Schiff, E. (1964). Further studies of overlearning and position reversal learning. *Psychological Reports, 14,* 380–382.

Davey, G. (1989). *Ecological learning theory.* Florence, KY: Taylor & Francis/Routledge.

Eibl-Eibesfeldt, I. (1970). *Ethology: The biology of behavior.* New York: Holt.

Fabricius, E. (1951). Zur Ethologie Junger Anatiden. *Acta Zoologica Fennica, 68,* 1–175.

Falk, J. L. (1961). Production of polydipsia in normal rats by an intermittent food schedule. *Science, 133,* 195–196.

Garcia, J., & Garcia y Robertson, R. (1985). Evolution of learning mechanisms. In B. L. Hammonds (Ed.), *Psychology and learning.* Washington, DC: American Psychological Association.

Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1974). Behavioral regulation of the milieu interne in man and rat. *Science, 185,* 824–831.

Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science, 122,* 157–158.

Garcia, J., & Koelling, R. A. (1966). The relation of cue to consequence in avoidance learning. *Psychonomic Science, 4,* 123–124.

Harlow, H. F. (1971). *Learning to love.* San Francisco: Albion.

Harlow, H. F., & Suomi, S. J. (1970). Nature of love—Simplified. *American Psychologist, 25,* 161–168.

Hess, E. H. (1962). Ethology: An approach toward the complete analysis of behavior. In R. Brown, E. Galanter, E. H. Hess, & G. Mandler (Eds.), *New directions in psychology* (pp. 157–266). New York: Holt.

Hess. E. H. (1964). Imprinting in birds. *Science, 146,* 1128–1139.

Hutchinson, R. R., Azrin, N. H., & Hunt, G. M. (1968). Attack produced by intermittent reinforcement of a concurrent operant response. *Journal of the Experimental Analysis of Behavior, 11,* 498–495.

Jacquet, Y. F. (1972). Schedule-induced licking during multiple schedules. *Journal of the Experimental Analysis of Behavior, 17,* 413–423.

Jaynes, J. (1956). Imprinting: The interaction of learned and innate behavior: I. Development and generalization. *Journal of Comparative and Physiological Psychology, 49,* 201–206.

King, G. D. (1974). Wheel running in the rat induced by a fixed-time presentation of water. *Animal Learning and Behavior, 2,* 325–328.

Klein, S. B., Domato, G. C., Hallstead, C., Stephens, I., & Mikulka, P. J. (1975). Acquisition of a conditioned aversion as a function of age and measurement technique. *Physiological Psychology, 3,* 379–384.

Klein, S. B., Freda, J. S., & Mikulka, P. J. (1985). The influence of a taste cue on an environmental aversion: Potentiation or overshadowing? *Psychological Record, 35,* 101–112.

Klopfer, P. H. (1971). Imprinting: Determining its perceptual basis in ducklings. *Journal of Comparative and Physiological Psychology, 75,* 378–385.

Klopfer, P. H., Adams, D. K., & Klopfer, M. S. (1964). Maternal "imprinting" in goats. *National Academy of Science Proceedings, 52,* 911–914.

Knutson, J. F., & Kleinknecht, R. A. (1970). Attack during differential reinforcement of low rate of responding. *Psychonomic Science, 19,* 289–290.

Kovach, J. K., & Hess, E. H. (1963). Imprinting: Effects of painful stimulation upon the following response. *Journal of Comparative and Physiological Psychology, 56,* 461–464.

Levitsky, D., & Collier, G. (1968). Schedule-induced wheel running. *Physiology and Behavior, 3,* 571–573.

Lorenz, K. (1957). The past twelve years in the comparative study of behavior. In C. H. Schiller (Ed.), *Instinctive behavior* (pp. 288–317). New York: International Universities Press. (Original work published 1952)

Pavlov, I. (1928). *Lectures on conditioned reflexes: Vol. 1: The higher nervous activity of animals* (H. Gantt, Trans.). London: Lawrence and Wishart.

Revusky, S., & Parker, L. A. (1976). Aversions to drinking out of a cup and to unflavored water produced by delayed sickness. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 342–353.

Riccio, D. C., & Haroutunian, V. (1977). Failure to learn in a taste-aversion paradigm: Associative or performance deficit? *Bulletin of the Psychonomic Society, 10,* 219–222.

Schulman, A. H., Hale, E. B., & Graves, H. B. (1970). Visual stimulus characteristics of initial approach response in chicks (*Gallus domesticus*). *Animal Behavior, 18,* 461–466.

Seligman, M. E. P. (1970). On the generality of laws of learning. *Psychological Review, 77,* 406–418.

Shanab, M. E., & Peterson, J. L. (1969). Polydipsia in the pigeon. *Psychonomic Science, 15,* 51–52.

Silva, F. J., Timberlake, W., & Gont, R. S. (1998). Spatiotemporal characteristics of serial CSs and their relation to search modes and response form. *Animal Learning and Behavior, 26,* 299–312.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* New York: Appleton-Century-Crofts.

Skinner, B. F. (1948). Superstition in the pigeon. *Journal of Experimental Psychology, 38,* 168–172.

Staddon, J. E. R., & Ayres, S. L. (1975). Sequential and temporal properties of behavior induced by a schedule of periodic food delivery. *Behavior, 54,* 26–49.

Staddon, J. E. R., & Simmelhag, V. L. (1971). The “Superstition” experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review, 78,* 3–43.

Timberlake, W. (2001). Motivated modes in behavior systems. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 155–209). Mahwah, NJ: Erlbaum.

Timberlake, W., & Lucas, G. A. (1989). Behavior systems and learning: From misbehavior to general principles. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theory: Instrumental conditioning theory and the impact of biological constraints on learning* (pp. 237–275). Hillsdale, NJ: Erlbaum.

Timberlake, W., Wahl, G., & King, D. (1982). Stimulus and response contingencies in the misbehavior of rats. *Journal of Experimental Psychology: Animal Behavior Processes, 8,* 62–85.

Weidman, U. (1956). Some experiments on the following and the flocking reaction of mallard ducklings. *British Journal of Animal Behavior, 4,* 78–79.

Wetherington, C. L. (1982). Is adjunctive behavior a third class of behavior? *Neuroscience and Biobehavioral Reviews, 6,* 329–350.

Wilcoxon, H. C., Dragoin, W. B., & Kral, P. A. (1971). Illness-induced aversions in rats and quail: Relative salience of visual and gustatory cues. *Science, 7,* 489–493.

31

Animal Cognition

Bradley R. Sturz
Armstrong Atlantic State University

Kent Bodily, Michelle Hernández, Kelly Schmidtke, and Jeffrey S. Katz
Auburn University

Animal cognition refers to the mechanisms by which nonhuman animals process information and is primarily concerned with understanding how animals learn, store, remember, and respond to information from the environment. Although similar in many respects to the present-day fields of animal learning, behavioral ecology, and ethology, animal cognition has emerged as a distinct area of study by adopting the information-processing perspective from human cognitive psychology. The information-processing perspective assumes that animals form and manipulate internal representations about their environment. As such processes afford flexibility and adaptability in behavior necessary for survival in an ever-changing environment, the information-processing perspective has proved useful in explaining and predicting numerous behavioral phenomena in animals.

Historically, animal cognition finds its roots in the American tradition of animal learning (e.g., Edward Lee Thorndike) and the European tradition of ethology (e.g., Konrad Lorenz, Niko Tinbergen). Each tradition differs in approach and method of investigation. Whereas ethologists often stress adaptive specializations and focus their studies in the context of an animal's natural environment, those studying animal learning often stress general processes and focus their studies in the context of a laboratory. Despite the rich historical roots in these traditions, animal cognition did not emerge as a distinct field of study until shortly after the cognitive revolution of the 1960s. It was then that the information-processing perspective was infused into the fields of animal learning and ethology. Since that time, animal cognition continues to mirror its human counterpart in focus and theory while maintaining an interest in learning and behavior. Over the years, these traditions have been integrated into a contemporary synthetic approach (Kamil, 1988; Shettleworth, 1998) that attempts to elucidate both specialized and general cognitive processes in the laboratory and field.

THEORY

Today, information-processing and evolutionary perspectives (influence of selective pressures in shaping cognitive mechanisms) serve as theoretical frameworks for research on animal cognition. From these frameworks, two main theoretical approaches for understanding psychological phenomena in animals have been developed: *modularity* and *general process*. One of many current challenges facing today's scientists is discovering procedures that can distinguish between these two approaches (but see Bitterman, 1965, 1975; Kamil, 1988).

Modularity

The modularity account of psychological phenomena proposes that mental processes are domain-specific mechanisms called *modules* (e.g., Fodor, 1983; Shettleworth, 1998). Overall, modularity proposes that animals utilize a number of these psychological mechanisms (modules) for the purposes of solving specific environmental problems. Each module processes a specific type of information. For example, just as an anatomical mechanism such as the eye is highly adept at processing a specific type of information (light), psychological mechanisms (modules) are adept at processing specific types of information. Given that species adapted to flourish within an ecological niche, their unique cognitive modules also may have evolved to solve environmental problems specific to that niche. An implication of a modularity account is that different species may perform qualitatively differently when engaging in analogous tasks or problems because of these unique cognitive mechanisms.

General Process

The general-process account of psychological phenomena proposes that mental processes, instead of being species-unique, are qualitatively similar across various species (e.g., Bitterman, 1960; Papini, 2002). Given that species developed specialized adaptations to flourish within unique ecological niches, similar cognitive mechanisms also may have evolved either through homology or homoplasy to solve common environmental problems. An implication of a general-process account is that evolutionarily divergent animals may perform qualitatively similarly when performing in analogous tasks or problems because of these similar cognitive mechanisms.

GENERAL METHODS

Although investigations of animal cognition also occur in an animal's natural environment, the present chapter focuses primarily on laboratory research. To study animal cognition in the laboratory, researchers precisely control experimental conditions in a host of procedures and tasks, and although apparatuses vary widely, the most common is the operant chamber. The operant chamber is a completely enclosed rectangular space that allows for the presentation of stimuli and recording of responses. Stimuli are shown on a computer monitor or behind translucent buttons located in the front panel of the chamber. A touch screen surrounding the monitor (e.g., like that of an ATM machine) or the buttons themselves detect responses to the stimuli. From these responses, researchers determine measurements such as accuracy and reaction time. Such precise control of an animal's environment allows researchers to rule out or control factors (e.g., past experience) that often confound experimental results from the natural environment.

A unifying theme of animal cognition research is that experimenters measure animal behavior during the learning of some task. Usually, researchers implement specific performance criteria for the task. The performance criteria ensure that an acceptable level of learning occurs. Upon reaching the performance criteria, researchers often introduce novel situations. By measuring accuracy and reaction time during these novel situations, researchers can infer *what* the animal is learning and *how* it is solving the task. In short, these novel tests under steady-state conditions reveal how the animal encoded and processed information.

AREAS OF INTEREST

Numerous areas of research are open to those interested in animal cognition, including (but not limited to) attention, perception, memory, spatial cognition, timing, counting, concept learning, problem solving, consciousness, emotion, foraging, theory of mind, communication, language, and tool use (for reviews see Roberts, 1998; Shettleworth, 1998; Wasserman & Zentall, 2006). Below we introduce several of these research areas and briefly discuss common methodologies and general results unique to each.

Memory

The current study of animal memory closely mimics that found in human cognition and places emphasis on how previous experiences are represented, stored, and remembered. Thus, like research on human memory, research on nonhuman memory also involves the development of tasks that assess the dynamic processes of encoding, retention, and retrieval. Encoding refers to the transformation of sensory stimuli (e.g., visual, auditory, olfactory, gustatory, and tactile) into mental representations. Retention refers to the storage of these representations, and retrieval refers to the reactivation or later access of these representations. Researchers often subdivide memory into various memory systems (or cognitive modules), including separate systems for working memory and long-term (reference) memory. Reference memory consists of declarative and procedural memory, and procedural memory consists of semantic and episodic memory. Procedural memory refers to memory for cognitive and behavioral skills, whereas declarative memory refers to memory for environmental information. Semantic memory refers to memory for general information about the environment, whereas episodic memory refers to memory for specific events (i.e., the what, where, and when of an event). Evidence supports the existence of all of these memory categories for animals, albeit the data are controversial in some cases (e.g., episodic memory).

Methods and Results

Common methodologies used to assess memory include delayed-matching-to-sample (DMTS) tasks, symbolic matching tasks, list-memory tasks, radial-arm mazes, and cache recovery tasks. To test working memory, researchers often utilize a DMTS task. In a DMTS task, a trial begins with the presentation of a sample (e.g., a blue circle). To ensure the subject encodes the sample, a response to the sample stimulus (e.g., a touch) occurs, after which the sample disappears. After an interval of time elapses (known as the *retention interval*), two stimuli are shown (known as *comparison stimuli*). One of the comparison stimuli is the same as the sample and the other is different. A correct choice response, a response to the comparison stimulus identical to the sample, generally results in a food reward. Experimenters often manipulate the length of the retention interval and the number of stimuli that serve as the sample. By then measuring the influence of these variables on animals' performance, researchers are able to determine both the duration and capacity of working memory.

Much like DMTS tasks, symbolic matching tasks involve the presentation of a sample and a choice among comparisons. The main difference between the two tasks is that in the symbolic task the sample and the correct comparison are not identical. Instead, the researcher establishes the samples and their correct comparisons arbitrarily. For example, an animal may be shown a red square as the sample, then a horizontal line and a vertical line as comparisons. The animal may then receive the reward only after selecting the horizontal line. Because the sample and comparisons are not identical, perceptual similarities such as hue or brightness cannot be the basis of a comparison choice; thus, correct responding on symbolic tasks suggests that animals memorize (reference memory) the individual samples and their correct comparisons.

List-memory tasks also begin with the presentation of a sample stimulus. After the observing response, the sample disappears. Instead of the presentation of the comparisons, as in a DMTS task, another sample is shown that also requires an observing response. After a number of sequential sample item presentations (i.e., the list), researchers introduce a retention interval followed by a single probe item. The probe item can be either identical to one of the list items or different. The animal then responds according to whether the probe item was presented in the list of sample items. For example, a pigeon may peck a right button if the probe item was presented in the list or a left button if the probe item was not presented in the list. As with DMTS and symbolic tasks, the list-memory task is useful in determining both memory duration and capacity. However, the list-memory task also has the distinct advantage of revealing changes in performance as a function of list-item position. For example, list-memory tasks have revealed primacy (better memory for items present early in the list) and recency effects (better memory for items present late in the list) with both pigeons and monkeys, which parallel results from human participants in similar tasks (Wright, 2006).

The radial arm maze is an apparatus that consists of a central starting position with an even number of arms emanating from its center. All of these arms are accessible from the central starting position. Before a trial, the experimenter places food at the end of the arms (and, at times, in either a fixed or random pattern). An animal (usually a rat) is put into the start position and then searches for the food rewards. When searching for the food, rats remember which arms they previously visited, as they rarely make a repeat visit to an arm (Olton & Samuelson, 1976). This result occurs even if the animal is taken from the apparatus for a fixed duration (retention interval) before it continues searching. Although such performance by rats may be the result of fixed patterns of searching (procedural memory), it also may be the result of memory for locations visited within the maze (declarative memory).

Researchers studying episodic memory in animals often use cache recovery tasks. A typical cache recovery task involves allowing an animal to store food at specific locations and then measuring search location accuracy on later retrieval of these items. Clayton and Dickinson (1998) initially found that scrub jays (a food-storing bird) can learn that the decay timeline (i.e., how long it takes a food item to become inedible) for wax worms is more rapid (< 124 hours) than that for peanuts. These birds then cached both wax worms and peanuts in visuospatially distinct locations and later recovered these items after 4 or 124 hours. After delays of 4 hours between caching and recovery, scrub jays searched more often at the worm cache locations, whereas after delays of 124 hours they searched more often at the peanut cache locations. Such results suggest that the scrub jays encoded location (where), content (what), and duration (when) of their cache sites.

Spatial Cognition

The study of spatial cognition focuses on how animals acquire, process, and store spatial information (Roberts, 1998). Such information is critical to the survival of mobile organisms because it allows for the learning of locations and permits navigation between locations. Although researchers routinely investigate several aspects of spatial cognition (e.g., orientation, spatial memory), the bulk of spatial cognition research emphasizes mechanisms of spatial coding of position. Two separate reference systems for spatial coding of position are egocentric and allocentric (for a review see Newcombe, 2002; however, see Wang & Spelke, 2002 for a view based on only an egocentric system). Researchers define these systems with respect to how spatial locations are referenced—either to the self (termed *egocentric*) or to the environment (termed *allocentric*). Based on these reference systems, evidence supports the use of several processes for locating objects in the environment such as path integration, beacon homing, and landmark-based navigation.

In the absence of landmarks, many animals are able to return to a location from which they start (e.g., their nest). Such a phenomenon, called *dead reckoning*, occurs by path integration—a process in which an estimation of position is updated via speed of movement, direction of movement, and elapsed time of movement (Etienne et al., 1998; Etienne & Jeffery, 2004). However, if a landmark is available, and located at or near a goal, some animals can plot a course directly to the environmental stimulus. This beacon homing process serves as a simple and effective way to locate places in the environment. If, however, a landmark is some distance from the goal, animals are able to rely on landmark-based navigation. Landmark-based navigation is the process of determining a position and orientation by using objects in the environment with known positions (Gallistel, 1990).

Methods and Results

Researchers interested in spatial cognition use several methodologies. The most common apparatuses include water mazes and open fields, and the common component to almost all spatial cognition tasks involves animals searching for a goal location. For example, in a standard Morris water maze task (Morris, 1981), researchers place animals (usually rats or mice) into an enclosed circular pool containing murky water. A small visible platform in the pool serves as the goal location. The amount of time to reach the platform is the measure of learning, and, not surprisingly, animals quickly learn to swim to the platform to escape the water. Across trials the platform progressively becomes invisible (by submerging it). Various cues placed around the walls enclosing the pool or on the inside walls of the pool itself serve as the only consistent source of information as to the location of the platform. Under these circumstances, animals continue to locate the invisible platform. Even in the absence of the platform, the animal spends the majority of search time near the absent platform's location. Locating the invisible platform and searching in the appropriate area for the absent platform provide evidence that the animal relies on the cues surrounding the pool to locate the goal. By then moving or removing these cues, the experimenter is able to determine which specific cue(s) the animal uses to locate the goal. Overall, experiments using the water maze show that animals rely on both local cues (stimuli close to the goal location) and global cues (stimuli at some distance to the goal location) but show preferential responding to closer and larger stimuli.

An open field also is routinely used in the study of spatial cognition and consists of a search space covered with a material (known as *substrate*) such as sand or sawdust in which an animal can search. Initially, researchers place food above this substrate at a specific distance and direction from one or more landmarks. Progressively, on subsequent trials, the researchers bury the food and the animal must search for the hidden food. Animals are adept at locating the food and continue to search at the correct location even on trials in which food is absent. Searching at the goal location despite the absence of food suggests that animals do not rely on visual or olfactory cues to locate the food. Instead, animals utilize information from the landmarks. Much like the water maze, open field tasks reveal which cues (landmarks) animals use to locate the goal but, importantly, also are ideal for determining how animals use these landmarks. For example, pigeons trained to find a goal in the presence of a single landmark will adjust their search behavior accordingly following shifts of the landmark to novel locations (Cheng, 1988, 1989). Such shifts suggest that animals encode both the distance and direction of the goal location from the landmark (Cheng & Spetch, 1998).

Concept Learning

The study of concept learning focuses on how animals learn to categorize stimuli and make judgments about relationships between stimuli. Concept learning research falls into three broad groupings: perceptual concepts, associative concepts, and abstract concepts. Perceptual concept learning involves sorting (or categorizing) stimuli (e.g., pictures of birds, flowers, people, trees) by stimulus similarity into appropriate categories (Herrnstein, Loveland, & Cable 1976; Huber & Aust, 2006). Associative concept learning involves sorting stimuli by a common response or outcome regardless of perceptual similarity into appropriate categories (e.g., Urcuioli, 2006; Vaughan, 1988). Abstract concept learning involves judging a relationship between stimuli by a rule (e.g., sameness, difference, addition, subtraction). Researchers consider a rule abstract if performance with novel stimuli during transfer tests is equivalent to performance with training stimuli. Although there is little dispute among theorists concerning the ability of various animals to learn associative and perceptual concepts, controversy remains concerning whether animals can learn abstract concepts because of alternative accounts (Katz, Wright, & Bodily, 2007). Nonetheless, evidence for all three types of concept learning occurs in species as diverse as bees, capuchin monkeys, dolphins, parrots, pigeons, rhesus monkeys, and sea lions. Such research also reveals that the number of items present during training (also known as set size) can be critical to concept learning (Shettleworth, 1998).

Methods and Results

Researchers use a variety of methods to test for concept learning, and two of the more common include sorting tasks and the same/different (S/D) task. Using a sorting task, Bhatt et al. (1988) allowed pigeons to see a single visual stimulus from one of four categories: car, cat, chair, or flower. The pigeons then selected one of four simultaneously available buttons that each corresponded to one of the four categories. After learning to categorize the initial training images, the researchers introduced novel images from each category. Pigeons correctly categorized the novel images better than chance, suggesting that they learned the perceptual concepts of car, cat, chair, and flower. Additional studies found that increasing the number of training images (set size) improved transfer performance (e.g., Wasserman & Bhatt, 1992). Presumably, correct performance increased because novel pictures became more similar to the training stimuli with the increase in the number of training stimuli.

In the S/D task, researchers present an animal with two items within a single trial. Sometimes the pair of items are the same (e.g., a baseball and a baseball) and other times they are different (e.g., an apple and a grape). Depending on trial type (i.e., same or different), the animal receives reward for making the appropriate response. For example, an animal might first see two pictures on a computer monitor. Next, a green square and blue diamond are made available simultaneously to the left and right of the pictures. A response to the green square results in reward when the pictures are the same, and a response to the blue diamond results in reward when the pictures are different. To test whether animals simply memorize the training item pairs, researchers introduce novel pairs of items (items different from those during training). If the animal discriminates novel pairs as well as the training pairs, it is taken as evidence for abstract concept learning. When the training set (e.g., 8 items) consists of a small number of training pairs (e.g., 64 pairs: 8 identical and 56 different) animals do not transfer to novel items, but when the training set becomes progressively larger (16, 32, 64, 128, 256) the level of transfer increases to the point where it is equivalent to training performance (i.e., full concept learning; Wright & Katz, 2006).

Although set size has similar effects on perceptual and abstract concept learning, the mechanisms that underlie these two types of concepts are different. Specifically, researchers claim the set size effect occurs because of the similarity between training and transfer stimuli for perceptual concept learning and by the formation of a rule between stimuli for abstract concept learning.

Timing and Counting

Timing refers to a process by which an animal tracks and marks how "much" of a time interval or duration passes. Counting refers to a process by which an animal tracks and marks how "many" of a quantity occurs. Although timing and counting may be separate mechanisms, both involve differentiating an amount, and this functional similarity suggests that these processes may result from a common mechanism.

Timing

There are at least two types of timing: circadian and interval. Circadian timing relies on a single pacemaker—a biological oscillator that pulses within the body and marks 24-hour durations. Circadian timing is involved in cyclical patterns of behavior such as sleep/wake cycles. These cycles are thought to result from the synchronization of the pacemaker with a reliable external cue such as the sun. Interval timing also involves a pacemaker to determine the amount of time that elapses between discrete events.

Methods and Results

Researchers use a variety of methodologies to assess timing. For example, circadian timing studies might present a signal, such as a light, at an irregular time. If a light is presented at the beginning of the night to a nocturnal animal, such as a hamster, it will rest longer into the following day. It is as if the presentation of light offset the pacemaker and shifted the entire cycle forward in time.

Investigations of interval timing may use conditional discrimination tasks. For example, rats in an operant chamber may first learn one response (e.g., depression of a lever) after the presentation of a two-second tone and another, different response (e.g., depression of a different lever) after the presentation of an eight-second tone. To assess the accuracy of an animal's ability to time, investigators present new durations of the tone. During these critical test trials, animals often respond to the lever associated with the two-second tone for shorter durations and the lever associated with the eight-second tone for longer durations. This reliable, differential responding suggests animals track the interval duration.

Another method to investigate interval timing uses a fixed-interval schedule, where a stimulus is present (e.g., a light or a tone) and food reward follows the first response emitted after a specific amount of time. For instance, a pigeon may receive reward for the first peck that occurs five seconds after presentation of a tone. Not surprisingly, responding increases in number as the interval progresses, suggesting that the animal is tracking the amount of time elapsed since the presentation of the stimulus. Another procedure, the peak procedure, modifies the fixed-interval procedure by inserting a trial with a longer duration that occurs in the absence of food reward. Here, animals tend to respond the most at the point in time at which food should occur, and responding decreases after that point. Such response patterns suggest that animals track the amount of time elapsed as they are able to determine when food should occur.

Counting

Recall that counting is a process by which an animal tracks and marks quantities. Gelman and Gallistel (1978) proposed five criteria necessary to claim that an animal is capable of counting. The one-to-one principle states that each item in an array should receive a distinct tag. For example, in English the tags are one, two, and three. The stable-order principle asserts that the tags must have a reliable order. One always precedes two and two always precedes three. The cardinal principle states that the last tag represents the total number of items in the array—in this case, three. The abstraction principle requires that counting not occur by rote memorization—any array with three items must be responded to as three (i.e., transfer). Finally, the order-irrelevance principle states that as long as each item has a single tag and tags follow a reliable order, which object is given which specific tag is irrelevant. Such strict criteria force researchers to develop precise procedures for training and testing counting behaviors.

Methods and Results

Methodologies to study counting can be similar to conditional discrimination tasks used to test timing with the exception that experiments use numerosities instead of intervals. For example, Brannon and Terrace (2000) used a task with rhesus monkeys that required correctly making a sequence of responses. Specifically, the researchers presented the monkeys with four colored boxes that each contained one to four items. The size of the items within the box and the location of the boxes varied to ensure that the monkeys attended to the number of items within each box and not the amount of space taken up by the items or the location of the boxes. The researchers required the monkeys to touch each box once (the one-to-one principle) in a reliable ascending order (the stable-order principle) and end on the highest quantity (the cardinal principle). When novel items were presented in the boxes, the monkeys performed above-chance accuracy, indicating the learning was not purely due to memorization (the abstraction principle).

Problem Solving

Problem solving refers to a process in which animals overcome obstacles in order to obtain an outcome (Köhler, 1927). Problem-solving research seeks to delineate tasks that animals are capable of solving and attempts to explain how the animals solve these tasks. The design of experimental problems demands great care and creativity from researchers and often results in novel and interesting methods for testing the problem-solving abilities of animals.

Problem solving involves several components: representation, learning, and choice. These components serve as the focus for much of the problem-solving research (Lovett, 2002). In short, representation refers to information encoded internally; learning refers to changes in behavior resulting from experience, and choice refers to the selection of specific behaviors or objects among a multitude of possibilities.

Methods and Results

Methodologies for studying problem solving are numerous and diverse. As the development and selection of problem-solving tasks vary widely due to researcher preference and species under study, we present a select few of the more well-known, problem-solving tasks in lieu of a list of common procedures.

Wolfgang Köhler, when studying chimpanzees, devised a now-famous problem commonly known as the “box and banana” problem. In this problem, he suspended a banana in the air so as to be unreachable and then scattered various objects including wooden boxes about the immediate area. The chimpanzees initially attempted to obtain the banana by unsuccessfully jumping for it. However, one chimpanzee, Sultan, began to look back and forth from the banana to the nearby boxes. All of a sudden, Sultan grabbed a box, put it under the banana, climbed atop the box, and grabbed the banana. As if through a flash of insight, problem solving was sudden and complete. However, despite observing the entire process, the other chimpanzees did not execute the task. Other chimps, although engaged in behaviors associated with components of the task, failed to link the separate skills successfully to obtain the banana. For example, one chimp found a box that was not near the banana, climbed on top of it, and stared at the distant banana.

In other studies, Köhler continued to present animals with the task of acquiring food out of their reach. As with the “box and banana” problem, objects such as sticks or boxes were within the animal’s immediate reach. Much like with the “box and banana” problem, the animals were unsuccessful in their initial attempts at reaching, stretching, grabbing, or jumping for the food. However, a few animals obtained the food by utilizing the available objects. Importantly, manipulation of the location of the object in the chimpanzees’ visual field increased their success rates. Specifically, objects that were at locations too far out of the visual field resulted in lower success rates. Köhler suggested that the changes in the environment influenced the animal’s representation of the problem and thus its ability to arrive at an effective solution. In other words, representational changes (changes to internal codes of information) resulted in higher or lower success rates. This line of research provided the foundation for research on problem representation and its importance in the problem-solving process.

Thorndike’s (1898, 1911, 1932) famous “puzzle box” studies involved placing a hungry cat in a box with food outside of its reach. The only way for the cat to obtain the food was to trip a lever, pull a string, or push a pole inside the box. These various behaviors served to open a door, allowing the cat access to the food. Thorndike noted that initial attempts to escape the box were sporadic and haphazard. In time, the cat performed the proper operation and escaped the box. Thorndike also noted that as the number of trials within the apparatus increased, the time to escape decreased; the animals’ behaviors became more direct at activating the mechanism responsible for opening the door. From these observations, Thorndike argued that problem solving was a gradual process. In his opinion, this “trial-and-error” learning was responsible for the production of behaviors that eventually resulted in solutions. Perhaps, most important, this line of research led Thorndike to develop one of the most basic principles of learning, the *Law of Effect*, which states that responses that result in satisfying consequences are strengthened.

In other studies investigating the learning component of problem solving, Robert Epstein and colleagues (1984, 1985) carefully controlled the prior experience of pigeons before testing them on an analog of the "box and banana" problem. Pigeons learned (a) to push a box toward a green dot at random locations, (b) to step up onto a box under a small facsimile of a banana hanging from the ceiling of the operant chamber, and (c) to not fly or jump at the banana. After learning these separate behaviors, a pigeon would enter the chamber with the banana hanging overhead and the box's location off to one side. However, the green dot would be absent. Stepping up on the box did not allow the pigeons to reach the banana (as the box was not under the banana). All three pigeons solved the problem within about a minute! Perhaps most important, pigeons with slight changes in training (e.g., pigeons that learned to push the box but not toward the green dot; pigeons that learned to peck the "banana" but not step up onto the box) failed to solve the problem. Epstein and his colleagues concluded that an ability to solve a problem depends largely upon specific prior learning experiences.

A series of laboratory studies with wild-captured New Caledonian crows nicely demonstrates the choice component in problem solving (Chappell & Kacelnik, 2002; Weir, Chappell, & Kacelnik, 2002). In these experiments, a small bucket with a wire handle (like a small tin paint can) containing meat was lowered into a plastic tube. The crows were unable to reach the meat or the wire handle with their beaks because the tube was too deep. However, the researchers placed two pieces of garden wire next to the tube: One piece of wire was straight and the other was bent into a hook. Impressively, numerous crows selected the hook-shaped wire, extended the hook into the tube, and retrieved the meat-filled bucket.

Tool Use

Although problem solving and tool use involve many of the same components, modification, in contrast to simple selection, is necessary for a behavior to be considered tool use. For example, in other experiments with the crows mentioned in the problem-solving section, some individuals actually bent a straight piece of wire into a hook and proceeded to retrieve the bucket from the tube (Kenward, Weir, Rutz, & Kacelnik, 2005). As another example, some chimpanzees in the wild remove a small branch from a tree, break off all smaller branches from it, and then insert the thin flexible branch into a termite nest. Chimpanzees eat termites, and not surprisingly, the termites readily cling to the branch. Thus, the chimpanzees need only to remove the branch from the termite nest to obtain the food.

In both of the cases just mentioned, the animal uses a modified object to achieve a result. However, many examples of tool use exist, especially in the wild, that are not as clear-cut. For example, chimpanzees use broken tree branches like a hammer to crack nuts. Crows drop nuts onto rocks (and pavement) to crack them. Sea otters swim on their backs and carry a flat stone on their stomachs. The stone serves as an anvil against which to beat clams and expose the meat. Some troops of capuchin monkeys beat clams against tree trunks. In all these examples, animals clearly use objects or structures in their environment to solve a problem, but the behaviors are not examples of tool use because the objects used as "tools" are not modified. These and other examples continue to challenge the way we think and talk about tool use in animals. It is important to remember that regardless of how we as humans define tool use, animals will continue to use objects in their environment, some of which they will modify, to obtain food or solve other problems.

Methods and Results

As with problem solving, methodologies used for studying tool use are quite diverse. However, one of the more common experimental tasks is the "trap tube task." The trap tube itself is a hollow, foot-long, cylindrical piece of transparent plastic about an inch in diameter with a small reservoir (the trap) located directly in the middle. The tube is often mounted horizontally on a foundation, and researchers place a food reward inside the tube next to the reservoir. In order to obtain the food, the animal (often a primate) must insert an available stick into the appropriate end of the tube or else the stick will push the food into the reservoir. In this task, the use of the stick itself to obtain the food falls into the gray area between problem solving and tool use but certainly does not entail modification. However, when given a bundle of sticks tied in the shape of an H, many animals will remove the unnecessary sticks and produce one stick to insert into the tube (Visalberghi, Fragaszy, & Savage-Rumbaugh, 1995).

APPLICATIONS

Although the study of animal cognition is important in its own right, it has many important applications. For example, research illuminating mechanisms underlying animal cognition are especially relevant to human education and mental health. Specifically, the study of animal cognition often results in comparative models of cognition. If normal cognitive functioning can be well understood in simpler cognitive systems, it is often possible to adapt this knowledge to more complex cognitive systems. This fundamental knowledge about normal cognitive functioning may permit identification of cognitive deficits and, more important, the potential sources of these deficits. As a result, fundamental knowledge about the nature of cognition also has potential for advancements in technological and methodological innovations for improving education, diagnosis, and treatment of learning disorders, mental disorders, brain-based injuries, and degenerative diseases. It may also provide researchers in other areas of scientific study with invaluable opportunities to explore associated neurological mechanisms and efficacy of pharmaceuticals.

SUMMARY

Emanating from the American tradition of animal learning and the European tradition of ethology, the study of animal cognition is a field focusing primarily on understanding how animals learn, store, and respond to environmental information. Animal cognition has emerged as a distinct field of study due to the infusion of the information-processing perspective from human cognitive psychology into the fields of animal learning and ethology, coupled with its focus on general and specialized psychological processes in the field and laboratory. This approach results in specialized methods and tasks for studying such cognitive phenomena as memory, spatial cognition, timing, counting, concept learning, problem solving, and tool use. The information gleaned from investigations into animal cognition has important implications for gaining fundamental knowledge about the nature of cognition and its potential for advancements in technological and methodological innovations, such as improving education and the diagnosis and treatment of learning disorders, mental disorders, brain-based injuries, and degenerative diseases.

REFERENCES AND FURTHER READINGS

Beck, B. B. (1980). *Animal tool behavior: The use and manufacture of tools by animals*. New York: Garland Press.

Bekoff, M., Allen, C., & Burghardt, G. M. (2002). *The cognitive animal*. Cambridge, MA: MIT Press.

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & Knauss, K. S. (1988). Conceptual behaviour in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 219–234.

Bitterman, M. E. (1960). Toward a comparative psychology of learning. *American Psychologist*, *15*, 704–712.

Bitterman, M. E. (1965). Phyletic differences in learning. *American Psychologist*, *20*, 396–410.

Bitterman, M. E. (1975). The comparative analysis of learning. *Science, 188*, 699–709.

Boysen, S. T., & Capaldi, E. J. (1993). *The development of numerical competence: Animal and human models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1–9 by rhesus macaques. *Journal of Experimental Psychology: Animal Behavior Processes, 26*, 31–49.

Brown, M. F., & Cook, R. G. (Eds.). (2006). *Animal spatial cognition: Comparative, neural, & computational approaches.* Retrieved November 1, 2007 from www.pigeon.psy.tufts.edu/asc/

Chappell, J., & Kacelnik, A. (2002). Tool selectivity in a non-primate, the New Caledonian crow. *Animal Cognition*, *5*, 71–78.

Cheng, K. (1988). Some psychophysics of the pigeon's use of landmarks. *Journal of Comparative Physiology A*, *162*, 815–826.

Cheng, K. (1989). The vector sum model of pigeon landmark use. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*, 366–375.

Cheng, K., & Spetch, M. L. (1998). Mechanisms of landmark use in mammals and birds. In S. Healy (Ed.), *Spatial representation in animals* (pp. 1–17). Oxford, UK: Oxford University Press.

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*, 272–274.

Cook, R. G. (Ed.). (2001). *Avian visual cognition*. Retrieved November 1, 2007 from www.pigeon.psy.tufts.edu/avc/

Epstein, R., Kirshnit, C. E., Lanza, R. P., & Rubin, L. C. (1984). "Insight" in the pigeon: Antecedents and determinants of an intelligent performance. *Nature*, *308*, 61–62.

Epstein, R. (1985). The spontaneous interconnection of three repertoires. *The Psychological Record*, *35*, 131–141.

Etienne, A. S., Berlie, J., Georgakopoulos, J., & Maurer, R. (1998). Role of dead reckoning in navigation. In S. Healy (Ed.), *Spatial representation in animals* (pp. 54–68). Oxford, UK: Oxford University Press.

Etienne, A. S., & Jeffrey, K. J. (2004). Path integration in mammals. *Hippocampus*, *14*, 180–192.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Healy, S. (1998). *Spatial representation in animals*. Oxford, UK: Oxford University Press.

Heyes, C., & Huber, L. (2000). *The evolution of cognition*. Cambridge, MA: MIT Press.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 285–302.

Huber, L., & Aust, U. (2006). A modified feature theory as an account of pigeon visual categorization. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative Cognition: Experimental explorations of animal intelligence* (pp. 325–342). New York: Oxford University Press.

Hulse, S. H., Fowler, H., & Honig, W. K. (1978). *Cognitive processes in animal behavior*. Oxford, UK: Lawrence Erlbaum Associates.

Kamil, A. C. (1988). A synthetic approach to the study of animal intelligence. In D. W. Leger (Ed.), *Nebraska Symposium on Motivation* (Vol. 35, pp. 257–308). Lincoln: University of Nebraska Press.

Katz, J. S., Wright, A. A., & Bodily, K. D. (2007). Issues in the comparative cognition of abstract-concept learning. *Comparative Cognition and Behavior Reviews, 2,* 79–92.

Kenward, B., Weir, A. A. S., Rutz, C., & Kacelnik, A. (2005).Tool manufacture by naïve juvenile crows. *Nature*, *433*, 121.

Köhler, W. (1927). *The mentality of apes* (2nd ed). New York: Harcourt, Brace & Co.

Lovett, M. C. (2002). Problem solving. In H. Pashler (Series Ed.) & D. Medin (Vol. Ed.), *Stevens' handbook of experimental psychology* (3rd ed.). *Vol. 2: Memory & cognitive processes*, (pp. 317–362). New York: John Wiley & Sons Inc.

Morris, R. G. M. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, *12*, 239–260.

Newcombe, N. S. (2002). Spatial cognition. In H. Pashler (Series Ed.) & D. Medin (Vol. Ed.), *Steven's handbook of experimental psychology* (3rd ed.), *Vol. 2: Memory and cognitive processes* (pp. 113–163). New York: Wiley & Sons.

Olton, D. S., & Samuelson, R. J. (1976). Remembrance of places passed: Spatial memory in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 97–116.

Papini, M. R. (2002). Pattern and process in the evolution of learning. *Psychological Review*, *109*, 186–201.

Roberts, W. A. (1998). *Principles of animal cognition*. Boston: McGraw-Hill.

Rogers, L. J., & Kaplan, G. (2003). *Comparative vertebrate cognition*. New York: Plenum Publishers.

Roitblat, H. R., & Meyer, J. (1995). *Comparative approaches to cognitive science*. Cambridge, MA: MIT Press.

Shettleworth, S. J. (1998). *Cognition, evolution, and behavior*. New York: Oxford University Press.

Terrace, H. S., & Metcalfe, J. (2005). *The missing link in cognition*. Oxford, UK: Oxford University Press.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplement, 2,* 1–109.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York: Macmillan.

Thorndike, E. L. (1932). *Fundamentals of learning*. New York: Teachers College, Columbia University.

Tomasello, M., & Call, J. (1997). *Primate cognition*. New York: Oxford University Press.

Urcuioli, P. J. (2006). Responses and acquired equivalence classes. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative Cognition: Experimental explorations of animal intelligence* (pp. 405–421). New York: Oxford University Press.

Vaughn, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *14,* 36–42.

Visalberghi, E., Fragaszy, D. M., & Savage-Rumbaugh, S. (1995). Performance in a tool-using task by common chimpanzees, bonobos, an orangutan, and capuchin monkeys. *Journal of Comparative Psychology*, *109,* 52–60.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences*, *6*, 376–382.

Wasserman, E. A., & Bhatt, R. S. (1992). Conceptualization of natural and artificial stimuli by pigeons. In W. K. Honig & J. G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 203–223). Hillsdale, NJ: Erlbaum.

Wasserman, E. A., & Zentall, T. R. (2006). *Comparative cognition: Experimental explorations of animal intelligence*. New York: Oxford University Press.

Weir, A. A. S., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science*, *297*, 981.

Wright, A. A. (2006). Memory processing. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 164–185). New York: Oxford University Press.

Wright, A. A., & Katz, J. S. (2006). Mechanisms of *same/different* concept learning in primates and avians. *Behavioural Processes*, *72,* 234–254.

32

Comparative Psychology

Alan M. Daniel and Mauricio R. Papini
Texas Christian University

Regardless of their area of expertise, scientists who approach the study of behavior in terms of adaptive function, evolutionary history, and developmental mechanisms can be considered comparative psychologists. Comparative psychology tries to bridge the gap between psychology's focus on behavior and biology's focus on evolution. The idea that evolution often starts with a functional change, such as a novel foraging behavior or mating strategy, and morphological changes then ensue has appealed to biologists since the 19th century. The preeminence of function over form is found, for example, in Lamarck's famous example of giraffes stretching their neck to reach higher foliage, thus resulting in the transmission of longer necks to their offspring. Although modern evolutionary theorists have largely abandoned this Lamarckian mechanism, known as *the inheritance of acquired traits*, the initiating role of function in evolution has remained central. This principle can be illustrated by a similarly famous example first used by Darwin: the beak of the Galapagos finches. The relation between population changes in beak shape and size and the availability of seeds has become a primary example of the action of Darwin's natural selection mechanism in wild populations. Thus, the foraging behavior of finches plays a determinant role in the direction of evolutionary change in beak morphology.

Early Greek and Roman philosophers advocated evolutionary-like views, including some resembling the modern notion of natural selection. For the most part, however, ancient views of nature were not clearly evolutionary, as far as is known today. Aristotle, for example, proposed that humans are the acme of the hierarchy of life because of their possession of rationality. The notion that humans stand apart from the rest of the natural world has persisted for cultural and religious reasons throughout the centuries.

After the European Renaissance, naturalistic views of behavior regained strength. Among these views were the influential mechanistic theories of animal behavior proposed by Gómez Pereira and René Descartes. These theories focused on the notion of reflex, a simple unit of behavior that requires an external stimulus, some processing at the level of the central nervous system, and a motor response output. This simple idea would eventually become the basis for early developments in the neurophysiology of behavior and learning during the second half of the 19th century. Simultaneously, Darwin's theory of evolution by natural selection provided a sound basis for the notion of continuity across species, both physical and behavioral. Darwin himself emphasized that rudimentary mental processes should be found in nonhuman animals, leading to the notion of mental continuity and opening the door to an exploration of the evolutionary roots of psychological processes. These are the main elements of the historical context in which comparative psychology emerged as a discipline.

Clifford L. Morgan is credited with writing the first general *Introduction to Comparative Psychology* in 1894, in which he proposed his influential canon based on the law of parsimony. In criticism of questionable anecdotal evidence for mental continuity, Morgan wrote, "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale." This statement suggests that of several equally effective accounts of some behavioral effects, the one based on the simplest psychological process should be selected. For example, a behavioral phenomenon should not be attributed to an emotional response when it can be completely explained in terms of simpler reflexes. Morgan's canon is aimed at preventing experimenter bias and anthropomorphic interpretations of behavioral data.

During the 20th century, a movement within the biological sciences gave rise to *ethology*, a discipline closely related to comparative psychology. Traditionally, ethologists focused on "innate" rather than acquired behaviors and examined behavior in naturalistic or seminaturalistic settings, rather than under the controlled conditions of the laboratory. These approaches tended to converge during the second half of the 20th century. Multidisciplinary integration is the main contemporary feature of comparative psychology. The goal of this chapter is to provide an overview of the methods and ideas that are likely to drive comparative psychology in the 21st century.

THEORY

Evolutionary biologists traditionally focus on cladistic relations between species to determine the phylogenetic history of specific traits. For psychologists, it is difficult to determine the origin of behavior because common ancestors are extinct and the fossil record does not preserve behavioral features, only some morphological characters. This means that comparative psychologists must examine evolutionary trends in terms of the behavior of living species. Evolutionary theory provides two major concepts to organize such research: clade and grade.

A *clade* is a set of species sharing a common ancestor. For example, chimps, gorillas, orangutans, and other apes are classified in the family Pongidae. The unique characters of this particular clade reflect evolutionary novelties and provide an opportunity to study how such novelties originate and spread across species. Alternatively, species from different clades may evolve common traits independently; theorists refer to these traits as *grades*, and they transcend phylogenetic relations. Encephalization in birds and mammals provides an example of the concept of grade. Both birds and mammals exhibit brain sizes that are typically about 10 or so times larger than the brain size known for reptiles. Although birds and mammals evolved from different reptilian groups, they both achieved a significant degree of encephalization independently. Therefore, birds and mammals share a grade in terms of relative brain size, even though they are classified as distantly related in taxonomic terms.

The complementary nature of these concepts can be illustrated with the example of aquatic mammals, such as cetaceans (whales, dolphins) and sirenians (manatees). Genetic evidence has revealed that the closest living relatives of the cetaceans are the artiodactyls (even-toed ungulates), so much so that some taxonomists have combined the two into a single clade: order Cetartiodactyla. Included in this clade are some purely land-dwelling species (cows, deer), some that are entirely aquatic (dolphins), and some land dwellers that spend a good portion of their lives in and around water (hippopotami). Presumably, cetacean ancestors resembled the partially aquatic hippopotamus as part of the transition to purely aquatic life from their land-dwelling ancestors. The same is probably true of the sirenians, whose closest extant relatives are the proboscideans (elephants). Any land-dwelling lineage adapting to an aquatic environment is likely to experience similar transitional stages. In this regard, cetaceans and sirenians share a grade of adaptation to aquatic life. Even though the species are distantly related and evolved under different ecological conditions, within-grade comparisons may be useful to understand the evolutionary trends required for such transitions independently of the details of phylogenetic history.

Evolutionary trends common between species often involve adaptations that allow the animal to specialize less (i.e., greater behavioral plasticity) and therefore exploit more resources in its environment. The fact that these traits make a species less specialized is likely to be the reason for their generality across species. Learning ability, for example, is an adaptation that dramatically increases plasticity, allowing the organism to adjust to environmental changes and predict events based on prior experience. Learning mechanisms tend to be highly conserved across many species. In contrast, adaptive changes in behavior that increase specialization are commonly achieved through adaptations in contextual variables (e.g., motor function, motivation, sensory processing). One hypothesis to explain these observations is that changes in a contextual variable can be more selective for the specific adaptation because they require the modification of only one or a few systems. In contrast, changes in learning mechanisms have the potential to change multiple behavioral systems, with some of those changes being maladaptive. In either case, by looking at grades of learning in (cladistically) primitive species, comparative psychologists are likely to discover how learning evolved in derived species. Comparisons across both grades and clades are vital to the study of comparative psychology.

Evolutionary Principles

Evolutionary principles guide comparative psychologists in the study of behavior. Among these principles, evolution by natural selection has provided an important theoretical guidance for behavioral studies. Three conditions must be met for evolution by natural selection to occur:

- Phenotypic variability within a population
- Heritability (phenotypic variability must be, at least in part, heritable)
- Differential reproductive success of alternative characters.

Natural selection can occur in several ways. Selective pressures can favor individuals with a given phenotype (e.g., long neck) over individuals with the alternative phenotype (e.g., short neck). Across generations, the population becomes dominated by individuals with the favored phenotype because the underlying genes spread more successfully in the population, while the alternative phenotype becomes less common. Cases in which selective pressures favor an extreme version of a trait are known as *directional selection*. Alternatively, individuals with longer necks and shorter necks may be able to exploit different resources, whereas those with intermediate features may suffer from greater competition. Cases in which both extreme versions of a trait hold a selective advantage are known as *disruptive selection*. Traits at the extremes of the distribution may also be selected against, favoring moderate expressions, a case known as *stabilizing selection*.

Selection can favor behavioral phenotypes in the same way it favors morphological characters. For example, artificial selection experiments with the fruit fly *Drosophila melanogaster* have shown that animals can be selected based upon their tendency to move to the top (negative geotaxis) or to the bottom (positive geotaxis) of a 16-unit T-maze. Such disruptive selective breeding produces strains of animals that express these extreme phenotypes. In naturalistic settings, researchers have observed selection in different species to work in similar ways to produce different behaviors, such as variations in courtship and mating behaviors, nest building, aggression, and other traits.

Presumably, there can be situations in which behavioral or morphological plasticity is favored by selective pressures. Selection for plasticity and learning ability is called the Baldwin effect. In the early years of the 20th century, James M. Baldwin proposed that an initial period of increased plasticity is required for abilities to arise. Once these abilities are established, they may become “instinctive” in subsequent generations because plasticity is no longer an important selected outcome. This loss of plasticity is called genetic assimilation. *Genetic assimilation* involves the evolution (via natural selection) of characters that were triggered by environmental pressures acting on a plastic phenotype in the ancestors, but that are organized internally in the descendants, independently of the presence of an environmental trigger. In this way, phenotypic plasticity can be an important factor driving evolutionary change.

For individual organisms, the ability to pass on their traits through reproduction is what defines their fitness. One way of determining an individual’s fitness is to observe how many of its total offspring, across its lifetime, survive to adulthood. This is known as *lifetime reproductive success* (LRS). Because these data are usually difficult to collect, especially in naturalistic settings, researchers sometimes use less comprehensive methods to estimate LRS. For example, looking at reproductive success over one or several breeding seasons, the number of successful copulations, defense of territory, or other similar indicators can provide estimates of LRS.

Comparative Approaches

Comparisons across extant species, though necessary, are not without difficulties. Comparative research may involve direct assessments of behavior in different species, or alternatively, extensive analysis of behavioral processes in a single species at a time. Although relatively few researchers dare to tackle the problems of developing long-term research programs involving two or more species, many model their research with a single species in such terms that comparisons are encouraged, even if only across laboratories. Thus, comparative research is actually more common than it may appear at first sight because most (but not all) research with nonhuman animals in psychology is ultimately concerned with human applications.

Comparative psychologists have followed a common trend shared with those who contribute to evolutionary biology in a more general sense. Evolutionary theories of behavior can adopt two complementary views, one based on adaptive function and the other based on phylogenetic history. The adaptive functional approach has several characteristics:

- It stresses the relation between behavior and ecology.
- It studies closely related species with contrasting ecology.
- It aims at determining the contribution of behavior to reproductive success.
- It deals with microevolutionary divergence in behavior.

Researchers have implemented the adaptive functional approach in a variety of areas such as in the study of flavor aversion learning in rodents (see Chapter 35). One study tested the hypothesis that rapid flavor-aversion learning is an adaptive trait for species that exhibit a generalized diet, but is less advantageous for species that specialize in a narrow set of foods. This specialist-generalist hypothesis was tested by inducing flavor-aversion learning in several species of rodents from the genus *Dipodomys* (kangaroo rats). Although *D. merriami* consumes grains, seeds, and other plant materials, *D. microps* specializes almost exclusively in the leaves of chenopod plants. As predicted, *D. merriami* (the generalist) acquired flavor aversions faster than did *D. microps* (the specialist). Researchers attributed the difference to a greater aversion to novel foods in the generalist than in the specialist, a phenomenon known as *neophobia*.

The adaptive significance approach is appropriate for characters that exhibit rapid evolutionary change. The morphological analogy would be the variations in beak size and shape exhibited by Galapagos finches. Other biological characters are more stable in evolutionary terms, thus this approach would not help determine how they evolved. For example, if a researcher is interested in studying the evolution of the avian feather, then looking at the Galapagos finches may not be particularly illuminating because this character is likely to be stable in all these species. A comparison of peripheral structures in birds (feathers), dinosaurs (some of which exhibit feather-like structures), and extant reptiles (scales) may be more illuminating. In such a comparison, the emphasis is on the evolutionary (or phylogenetic) history of this particular character. When applied to behavior, this approach has the following characteristics:

- It stresses the relation between behavior and phylogenetic history.
- It studies distantly related species.
- It aims at determining homology, homoplasy, and divergence in mechanisms.
- It deals with macroevolutionary trends in behavior.

Most comparative research in psychology involves a phylogenetic approach. For example, during the past few decades, there have been numerous attempts at demonstrating that nonhuman animals ranging from apes and rats to parrots and pigeons have the potential to exhibit higher cognitive functions, including abstract concept formation, self-awareness, and language. Traditional comparative studies have also emphasized the generality of basic learning processes by studying Pavlovian and instrumental conditioning in simple organisms, such as the marine slug *Aplysia californica*, the nematode *Caenorhabditis elegans*, or the honeybee *Apis mellifera*. Systematic research with a variety of vertebrate species has also shown evidence of divergence in basic reinforcement processes, as shown by the study of a phenomenon known as *successive negative contrast* (SNC). In SNC, the experimenter trains a group of animals with a large incentive that is eventually downshifted to a small incentive. The behavior of these animals deteriorates beyond the level of a control group always given access to the small incentive. SNC and related effects occur in several mammalian species, including primitive marsupials like the opossum, to rats, pigs, monkeys, apes, and humans. However, analogous experiments with bony fish, amphibians, reptiles, and birds have failed to yield evidence of SNC. What these experiments indicate is that nonmammalian vertebrates perceive the downshift in incentive magnitude, but show no sharp deterioration in behavior. Researchers have speculated that SNC depends on mechanisms of emotional arousal (e.g., frustration) that are present only in adult mammals.

Methodological Issues

Comparative studies face complex methodological problems that can be illustrated with the following experiment. Imagine that a researcher wants to establish a direct comparison between the rates of learning in rats and pigeons. The researcher matches training parameters as much as possible by administering practice in the same apparatus, under the same trial-distribution conditions, with the same type and amount of food, using adult animals, and so on. If the species differ in acquisition rates, is it fair to conclude that this demonstrates species differences in learning processes?

The short answer is "no." The reason for this negative answer lies in the impossibility to demonstrate that objectively equal conditions have the same impact on the behavior of both species. For example, consider the amount of food given as the incentive during training trials, although the same logic applies to all other aspects of the training situation. The same amount may be more rewarding to the rat than to the pigeon (or vice versa). Thus, differences in performance across species may reflect the differential effects of motivational, perceptual, or motor processes, rather than learning mechanisms per se. These factors are called *contextual variables*. Because of species differences in contextual variables and the impossibility of equating them, researchers have turned away from direct comparisons across species.

The alternative methodology involves an emphasis on the comparative study of functional relations between variables that can affect learning and behavioral outcomes. Thus, a fruitful approach would be to ask whether relatively larger reinforcers yield faster learning in both rats and pigeons, because absolute comparisons can be greatly affected by contextual variables rather than by learning mechanisms. When different species show different behavioral outcomes under analogous variations in conditions of training, the question arises as to whether the mechanisms of learning have diverged. As in other types of research, comparative psychologists studying behavioral plasticity, whether in the context of learning, cognition, or specific behavioral phenomena (e.g., play behavior, courtship, parental care, etc.), frame their findings in terms of the theoretical concepts provided by evolutionary principles, including homology, homoplasy, divergence, adaptation, and so on.

Research in comparative psychology often aims at identifying the mechanisms underlying specific psychological effects. This approach implies the need for a clear definition of "mechanism," a concept that has acquired different meanings for different researchers. The need to characterize this concept is also highlighted by comparative research aimed at determining whether analogous behavioral outcomes are homologous or homoplasic across species. *Homology* implies inheritance from a common ancestor, whereas *homoplasy* refers to the convergence of behavioral outcomes, despite independent evolution, because of common ecological pressures.

A useful method for determining whether species share homology of behavioral mechanisms is to study the phenomenon at different levels of analysis. Behavior can be understood at at least four levels:

- Psychological level (e.g., associations, fear, timing clock)
- Neurobiological level (e.g., neural networks in specific brain sites)
- Neurochemical level (e.g., neurotransmitter systems)
- Cell-molecular level (e.g., second-messenger systems, gene expression).

Lower levels are more basic, and therefore tend to be more general across species and behaviors. As the hierarchy of levels progresses, the focus becomes more and more specific to a particular behavioral phenomenon. Across species, the same behavioral output may be achieved from the same mechanisms (homology) or different mechanisms (homoplasy) at any of these levels. For example, a comparison of a common behavioral outcome in two species that yields evidence of similar mechanisms at all four levels supports the hypothesis that the underlying mechanisms are homologous. If, however, species differ in terms of mechanisms at some or all levels, but the behavioral outcomes are similar, this result is consistent with the homoplasy of underlying mechanisms.

Recent research in several areas shows a third alternative. For example, mollusks, insects, and rodents have a very remote common ancestor, an animal that lived more than 600 million years ago and was no more complex in neural terms than a planarian. Despite the independent evolution of the organization of the central nervous system, these animals share some basic cell-molecular mechanisms of learning. In general, when cellular and genetic processes are homologous, but the higher-level character has evolved independently, biologists invoke the concept of *parallel evolution*. In parallel evolution, the sharing of common building blocks increases the likelihood that independently evolved traits will have similar morphology or function.

APPLICATIONS

The main contribution of comparative psychology relates to a better understanding of the processes that lead to the evolution and development of animal behavior. Traditionally, comparative studies have been treated as basic research carried out mainly for the purpose of increasing human understanding of nature. Although researchers have focused on behavior, these studies have had implications for an understanding of other problems, including brain evolution, human evolution, economic theory, neural networks, and others.

Inevitably, however, basic research leads to applications aimed at solving specific problems of human relevance. Some of the areas to which comparative psychologists have contributed include the following, in no special order:

- Development of psychoactive drugs
- Animal models for neurological disorders
- Improving animal production
- Environmental enrichment for zoo animals
- Development of psychotherapeutic techniques applied to mental disorders
- Animal training for law enforcement and the TV and film industries
- Improving survival and reproductive success of endangered species
- Determining the behavioral consequences of pollution in natural populations
- Restricting the interaction between wild predators and cattle

SUMMARY

Comparative psychology is among the oldest branches of scientific psychology. Well over one century of behavioral research has led to an impressive improvement in an understanding of the evolutionary and developmental basis of behavior. A century ago educated people had little or no understanding of the perceptual capacity of animals, their learning and memory skills, their ability to make and use tools, their higher cognitive powers, and the extent to which their behavior was an expression of neural activity, to name but a few categories in which significant progress has been made. An intended consequence of these studies has been the assumption that they will provide a better framework to understand human nature. By appreciating the psychological skills of nonhuman species, comparative psychologists have come a long way in their contribution to identifying the characteristics that humans share with other species, as well as those that may be uniquely human—the natural product of a long evolutionary history.

REFERENCES AND FURTHER READINGS

Bitterman, M. E. (1975). The comparative analysis of learning. *Science, 188*, 699–709.

Bitterman, M. E. (1996). Comparative analysis of learning in honeybees. *Animal Learning and Behavior, 24*, 123–141.

Burns Cusato, M. (Ed.). (2002). Comparative psychology and the applied challenge [Special issue]. *International Journal of Comparative Psychology, 15,* 2–3.

Burns Cusato, M. (Ed.). (2003). Comparative psychology and the conservation of endangered species [Special issue]. *International Journal of Comparative Psychology, 16,* 2–3.

Gottlieb, G. (1992). *Individual development and evolution: The genesis of novel behavior*. Oxford, UK: Oxford University Press.

Grant, P. R. (1985). *Ecology and evolution of Darwin's finches*. Princeton, NJ: Princeton University Press.

Heyes, C., & Huber, L. (Eds.). (2000). *The evolution of cognition*. Cambridge, MA: MIT Press.

Karin-D'Arcy, M. R. (2005). The modern role of Morgan's canon in comparative psychology. *International Journal of Comparative Psychology, 18*, 179–201.

Morgan, C. L. (1895). *An introduction to comparative psychology*. London: Walter Scott.

Papini, M. R. (2002). *Comparative psychology: Evolution and development of behavior*. Upper Saddle River, NJ: Prentice Hall.

Papini, M. R. (2003). Comparative psychology of surprising nonreward. *Brain, Behavior and Evolution, 62*, 83–95.

Ricker, J. P., & Hirsch, J. (1985). Evolution of an instinct under long-term divergent selection for geotaxis in domesticated populations of *Drosophila melanogaster. Journal of Comparative Psychology, 99*, 380–390.

Shettleworth, S. J. (1998). *Cognition, evolution, and behavior*. Oxford, UK: Oxford University Press.

Striedter, G. F. (2005). *Principles of brain evolution*. Sunderland, MA: Sinauer.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford, UK: Oxford University Press.

PART VI

BASIC LEARNING PROCESSES

33

Classical Conditioning

Martha Escobar

Auburn University

If you touch a doorknob and receive an electric shock, you will likely hesitate to touch the doorknob again. If you eat an exotic food and later feel sick to your stomach, you will likely avoid consuming that food in the future. You use your previous experience with the world to determine your future behavior. This capacity is partially due to the process of *classical conditioning*.

THE DISCOVERY OF CLASSICAL CONDITIONING

Classical conditioning was first extensively studied by Ivan Pavlov (1927) in the early 20th century. Pavlov was a Russian physiologist interested in the processes of digestion, specifically the production of saliva and gastric juices that result from tasting food. His subjects (dogs) were restrained and meat powder was placed on their tongue; the saliva and gastric juices that resulted from their tasting the food were then collected and measured using special devices. Salivation upon tasting food is known as a *reflex* because salivation is involuntarily elicited by the experience of food. Reflexes were thought to be innate and impossible to change with experience. However, Pavlov observed that the sight of an experimenter and the sounds of equipment calibration were enough to make the dogs salivate. That is, events that preceded the delivery of the meat powder were also sufficient to elicit the salivation reflex. Pavlov and his students were amazed by this discovery. Their dogs were able to anticipate what would occur in the near future! To make sure their observations were real, they designed an experiment in which the sound of a metronome preceded the delivery of meat powder to the dogs' tongue. After many repetitions of these metronome-meat powder pairings, dogs were presented with the sounds of the metronome but no meat powder was delivered to their tongue. Because the dogs salivated to the sound of the metronome, Pavlov suggested that salivation, a reflexive *response*, could be produced by two different *stimuli*: those that produced salivation without prior experience (food), and those that subjects experienced as predicting the delivery of food (metronome).

THE ELEMENTS OF CLASSICAL CONDITIONING: US, UR, CS, AND CR

When we talk about classical conditioning, we talk about *stimuli* producing *responses*. (Note that *stimuli* is the plural for *stimulus*.) A *stimulus* is an event that subjects can perceive through their senses. Thus, lights, sounds, vibrations, odors, tastes, temperature, pain, and internal sensations (e.g., stomach rumbling) can all be considered stimuli. *Responses* are reactions to stimuli that are usually described in terms of behavior. Thus, salivation, shivering, jumping, walking, and other behaviors are responses.

Classical conditioning occurs when two stimuli are paired with each other. Usually, one of those stimuli

Before conditioning	*During conditioning*	*After conditioning*
Metronome = No response NS **Meat powder** = **Salivation** US = UR	Metronome → **Meat powder** = **Salivation** NS → US = UR With repeated pairings, NS becomes CS	Metronome = Salivation CS = CR

Figure 33.1 Pavlov's basic procedure.

produces a response innately (i.e., without experience). Pavlov called these stimuli *unconditioned stimuli* (USs) because the response they elicit is not conditional on things other than the US occurring. The response produced by a US is known as an *unconditioned response* (UR), and it usually entails a reflex elicited by the US (e.g., salivation, muscle flexion, etc.). In Pavlov's experiment with dogs, the stimulus that produced the salivation response unconditionally was meat powder. Thus, meat powder is a US that produces a salivation UR. In that experiment, the sound of a metronome, a neutral stimulus (NS), which did not elicit a response, preceded the presentation of meat powder. However, with repeated metronome-meat powder pairings, the metronome became a *conditioned stimulus* (CS), which produced a *conditioned response* (CR) of salivation. The salivation produced by the CS is known as conditioned because its occurrence is conditional on the CS and US having been previously paired; that is, because it results from the process of conditioning.

LABORATORY METHODS USED TO STUDY CLASSICAL CONDITIONING

Conditioned Emotional Response (CER)

When rodents (e.g., rats and mice) anticipate danger (e.g., an approaching predator), they produce a very specific behavior: they *freeze*. Freezing involves becoming very still to avoid detection by predators. Estes and Skinner (1941) developed the conditioned emotional response (CER) procedure to take advantage of this tendency to freeze in anticipation of danger. In a CER study, subjects are first trained to perform a behavior such as pressing a lever to receive food reward. Once this behavior is well established, the CS-US pairings begin. The nature of the CS may vary (lights, tones, etc.), but the US is always a stimulus that triggers fear reactions, commonly a mild electric shock (just strong enough to startle the subject without producing pain). Because the CS signals the occurrence of the US, subjects come to freeze when presented with the CS. Freezing is incompatible with performing behaviors such as lever pressing; thus, disruption of this behavior by freezing can be used as a measure of conditioned fear. Because conditioned fear suppresses the occurrence of the behavior, the measure is known as *conditioned suppression*. Suppression scores are sometimes confusing because less behavior indicates more learning about the CS-US relation. Keep in mind that in the CER procedure we do not measure the amount of behavior, but rather the suppression of that behavior. Thus, more learning would be indicated by more suppression (less behavior) and less learning would be indicated by less suppression (more behavior).

Sign Tracking

All animals tend to approach and make contact with stimuli associated to food (they *track* the food *signs*). Birds have a stereotypical tendency to peck at stimuli that signal food. In the sign-tracking preparation (Hearst, 1975), researchers repeatedly present pigeons with a CS (e.g., a lighted button) that signals delivery of a food US (grain). At first, pigeons eat the grain and ignore the CS. However, with repeated pairings, they begin to attend to the CS, and eventually come to peck at it as soon as it is presented. Thus, the researcher can record the number of pecks at the CS to measure learning of the CS-US relation. The stronger the CS-US relation, the more they will expect the US to follow the CS, and the more they will peck at the CS.

Conditioned Taste Aversion (CTA)

Most (if not all) animals tend to avoid foods that made them ill in the past (see Chapter 35, Taste-Aversion Learning). In the conditioned taste aversion (CTA) preparation, rats are presented with a novel flavor CS (e.g., sugary water) and then they are made ill to their stomachs by either injecting them with a drug or exposing them to an illness-inducing US (e.g., radiation). As a result of these CS-US pairings, subjects come to anticipate becoming ill after consuming the CS and avoid consuming it. Experimenters measure either total consumption of the CS (more learning about the CS-US relation results in less consumption of the CS) or preference for the CS over the other flavor (more learning about the CS-US relation results in higher preference for the other flavor). Food preferences develop through a similar process, but in this case the CS is paired with an *appetitive* US and consumption of the flavor CS increases (see Sclafani, 1997).

Human Conditioning

Diverse methods have been used for studying conditioning with humans, ranging from delivering mild electric shock to participants' fingertips and measuring finger flexion, to use of video games that allow experimenters to observe online records of participants' behaviors. The tasks are very diverse, but they all present participants with pairings of a CS and a US and require subjects to provide a measurable response.

CSs or USs? A Revised Definition

Is food a CS or a US? In the sign-tracking preparation, food is a US that, when paired with a visual CS, elicits a CR of pecking. However, in the CTA preparation, food is a CS that, when paired with an illness-inducing US, produces avoidance of a flavor. The terms CS and US are not absolute; they are relative to the role that the stimulus plays in the situation. Food is a US in the sign-tracking preparation because it innately elicits a pecking UR. In contrast, food is a CS in the CTA preparation because avoidance is not elicited by the food CS without learning. Thus, a stimulus should not be labeled "CS" or "US." These labels depend on the preparation and the responses being measured (e.g., Gunther, Miller, & Matute, 1995).

Do All CRs Mimic the URs?

In the sign-tracking preparation, the CR (pecking at the CS) closely resembles the UR (pecking at the food US). However, this is not always the case. In the CER procedure, the UR to shock is an increased level of activity (think about your own reactions to receiving an electrical shock), but the conditioned response is a decreased level of activity (freezing). Thus, CRs and URs are sometimes identical, but they can also be quite different from each other.

Is the CR Elicited Only by the CS?

Sometimes, stimuli other than the CS can produce the CR. This effect is known as *stimulus generalization* and it often occurs when the stimulus being tested and the CS are similar (Pavlov, 1927). For example, a pigeon trained to peck at a red circle may also peck at a highly similar orange circle, thus generalizing the CR. The opposite process is also observed: the same pigeon would not peck at a blue circle because the similarity between red and blue is low; that is, the pigeon would exhibit *stimulus discrimination*.

ACQUISITION OF CONDITIONED RESPONSES

The process of developing a CR as a result of CS-US pairings is known as *acquisition*. Acquisition can be favored and hampered by several manipulations. First I review the conditions that determine the strength of the acquired CR, and then I describe manipulations that diminish or prevent the development of a CR.

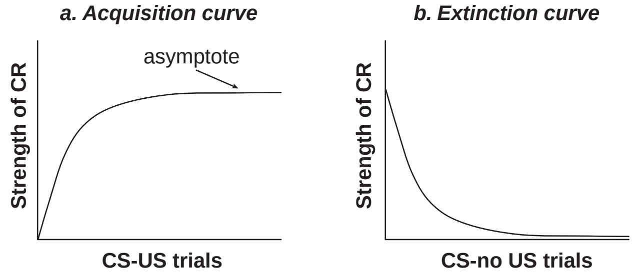

Figure 33.2 Acquisition curve and extinction curve.

NOTES: (a) Speed of acquisition of the CR slows down as the number of CS-US pairings increases, until asymptote is reached. (b) Extinction curves are also negatively accelerated, with the magnitude of change in CR strength decreasing as the number of CS-no US trials increases.

Number and Distribution of CS-US Pairings

In general terms, the more times the CS and US are paired, the stronger the CR that is observed. But not all pairings are equal. Most of the behavior change observed as a result of conditioning (i.e., acquisition) occurs during the first few trials. With each passing trial, the change in behavior is less and less. Plotted graphically, we would see a *negatively accelerated curve*, which means that the gradient of the curve is steeper during the first few trials, becoming less steep as training progresses, and finally becoming almost flat at a point known as the *asymptote of learning*. Asymptotic behavior represents the maximum strength of CR possible with that particular conditioning preparation. Figure 33.2 presents a hypothetical acquisition curve.

The steepness of the acquisition curve and the level of the asymptote are affected by the *salience* (i.e., perceived intensity) of the CS and US: More salient stimuli result in faster learning (steeper curve) and more CR (higher asymptote; see Kamin, 1965). Thus, you will learn faster to fear a loud siren signaling a dangerous fire than a dull tapping sound signaling an uncomfortable temperature. Importantly, salience is a *perceived intensity*, meaning that some CSs and USs may be more or less salient depending on the subject's particular state (e.g., food is more salient to a hungry subject than to a satiated one).

Each CS-US pairing is known as a *trial*, and the interval between trials where there are no CS or US presentations is known as the *intertrial interval*. In the short run, CRs are stronger with shorter intertrial intervals (this arrangement is known as "massed" trials). However, these CRs are weak and tend to decrease as time elapses after the last CS-US trial. To observe CRs that will stand the passage of time, the CS-US pairings must be given with long intertrial intervals ("spaced" trials). This is why your professors recommend that you study every day and review the material over and over again rather than cram the night before an exam: Cramming may help you pass an exam the following day, but the material will be lost by the time finals roll in.

Contiguity

Imagine that you open a door by turning a doorknob and 10 seconds later you feel an electric shock in your hand. Imagine now that touching the doorknob is immediately followed by receiving an electric shock. In which of these two situations would you be more likely to avoid touching doorknobs in the future? Which situation makes you more certain that the doorknob signaled the shock? These two situations have the same CS (doorknob) and US (static electricity shock) pairings, but the time elapsed between CS presentation and US delivery is not the same. That is, these two situations vary in the degree of *contiguity* between the CS and US. Conditioning appears to be stronger the greater the degree of CS-US contiguity: As contiguity degrades, it appears more difficult for subjects to form CS-US associations (Rescorla & Cunningham, 1979).

The most common CS-US arrangement is the one in which the CS precedes the US by a short time; the US either occurs immediately following termination of the CS or at the time of CS termination. This is known as *delayed conditioning* because the delivery of the US is delayed until the end of the CS. CRs resulting from delayed conditioning tend to be very robust, but the strength of these CRs decreases as the interval between CS termination and US delivery is increased. At this increased interval, we talk about *trace conditioning* because the US presentation occurs during a memory trace of the CS. As the CS termination-US delivery interval grows longer, conditioning becomes less robust. If you were to see a bee standing on your arm and immediately feel a painful sting (a delayed CS [bee]—US [sting] pairing), you would be more likely to attribute the painful sting to the bee and fear bees in the future than if you see a bee standing on your arm and feel a painful sting several seconds later (a trace CS-US pairing).

Contiguity does not imply that the CS precedes the US. Indeed, in some situations the US may occur before the CS. Because this is a reversal of the typical conditioning situation, US-CS pairings are known as *backward conditioning*. Continuing with the previous example, if you were to feel a painful sting and then saw a bee standing on your arm, you would be subject to backward conditioning. Backward conditioning sometimes results in little CR; indeed, subjects may view the CS as a signal that the US is already over.

But maximal contiguity does not necessarily mean maximal responding. The greatest possible CS-US contiguity is provided by situations in which the CS and US are presented at the same time, something known as *simultaneous conditioning*. This type of conditioning usually results in little CR when measures such as freezing (which require subjects to prepare for a US that will soon occur) are used: if the US is already present, producing an anticipatory response may not be advantageous to subjects (Matzel, Held, & Miller, 1988).

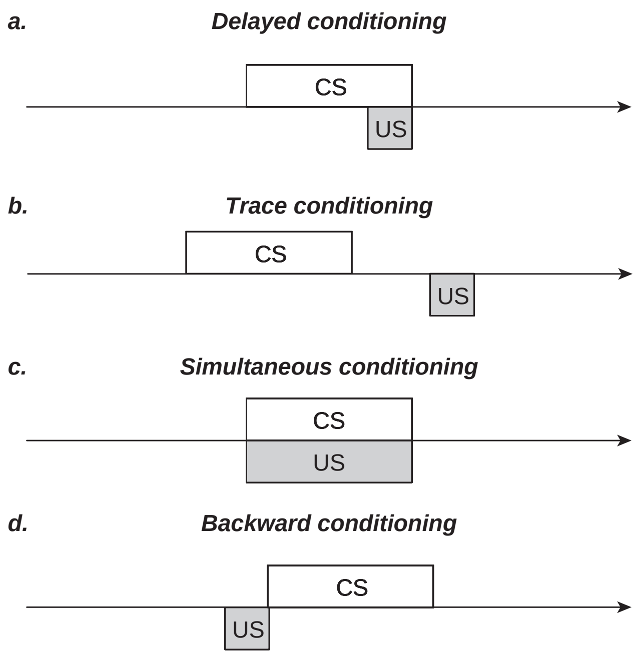

Figure 33.3 Common classical conditioning procedures.

NOTE: The arrow represents time elapsing. The width of the rectangles represents stimulus duration. *a. Delayed conditioning:* US delivery is delayed until CS termination. *b. Trace conditioning:* There is a short gap between CS termination and US delivery. *c. Simultaneous conditioning:* The CS and US occur at the same time. *d. Backward conditioning:* The US occurs before the CS (a reversal of the usual CS→US arrangement). The four procedures are presented in the usual order of effectiveness in producing CRs, from most effective (delayed conditioning) to least effective (backward conditioning).

Contingency

Paul develops heartburn every time a spicy meal is consumed, whereas Justin develops heartburn only occasionally when consuming spicy food. Which of these men is more likely to select a spicy meal from a restaurant menu? Probably Justin. In both cases, the *contiguity* between the CS (spicy food) and the US (developing heartburn) is the same: Heartburn occurs a few hours after consuming the spicy food. However, the contingency between these events is not the same. The relation is more reliable (i.e., the contingency is greater) in Paul's case than in Justin's case.

Contingency refers to the frequency with which the CS and US occur together versus apart (e.g., Rescorla, 1968).

Contingencies can be positive, negative, or zero. There is a positive contingency between Paul's eating spicy food (CS) and his developing heartburn (US; the CS and US occur more often together than apart). In contrast, there is a negative contingency between Paul's taking an antacid (CS) and his developing heartburn (US; the CS and US occur more often apart than together). Finally, there may be a zero contingency between Paul's eating a steak (CS) and his developing heartburn (US; the CS and US are equally likely to occur together or apart).

Note that contiguity and contingency are related to each other. For example, as contiguity degrades, contingencies start becoming less positive. If the CS and US occur close together *and* they only occur if the other is presented, both contiguity and contingency will be maximal and stronger CRs will be observed.

ELIMINATING CONDITIONED RESPONSES: EXTINCTION

On one occasion, my dog was playing in the backyard with his favorite toy, a red ball, when a construction truck down the street dropped something, scaring him badly. Afterward, he was afraid of his ball. What could I do to make my dog overcome this unreasonable fear? Let's take a look at the learning situation: A CS (ball) was paired with a US (loud noise). If, as a result of classical conditioning, my dog learned that the CS signaled the occurrence of the fear-producing US, maybe the CR (fear) can be attenuated by teaching him that the CS is *not* a signal for US occurrence. That is, if he was to experience the CS *without the US*, then he could come to expect no US to follow the CS. After allowing my dog to play with his ball for a few days without any loud-noise incidents, he moved from reluctantly approaching the ball (strong fear CR) to happily playing with it (weak or no fear CR).

Presenting the CS without the US (i.e., CS-noUS) following acquisition of the CR through CS-US pairings is known as *extinction*. Extinction *is* a type of learning; the subject is learning that the CS no longer signals the US. Indeed, extinction and acquisition produce changes in behavior that are very similar, although opposite in direction. Just as with acquisition, most of the extinction we observe with CS-noUS training occurs during the first few trials. Thus, extinction curves are also *negatively accelerated*, but they are the mirror image of acquisition curves because CRs are decreasing rather than increasing (see Figure 33.2).

Extinction eliminates a CR, but it should not be confused with eliminating learning. Sometimes, the CR may return after extinction has taken place. For example, I had some work done on my backyard and my dog could not play with his ball for a few weeks. When I presented him with the ball again, he was somewhat reluctant to approach it and play with it. That is, somehow, without further conditioning, the fear CR that had undergone extinction returned. This phenomenon is known as *spontaneous recovery*, and it can be defined as a dissipation of extinction due to the passage of time (Pavlov, 1927). But time is not the only way in which extinguished responses "recover," or return after extinction. Sometimes, returning to the environment in which the CS-US association was acquired is enough to bring the CR back (environmental stimuli are usually known as *contexts*). For example, my dog acquired fear of his ball in my backyard (Context 1). If I had conducted the extinction treatment in the park (Context 2) until he happily played with his ball, I might have seen some fear CRs when presenting him with the ball in my backyard (Context 1). Thus, even though a CR has undergone extinction, it could return if the subject is taken back to the context in which the response was acquired. This return of a response due to exposure to the CS in a context other than the extinction context is known as *renewal* (Bouton & Bolles, 1979)

TO RESPOND OR NOT TO RESPOND: EXCITATION AND INHIBITION

So far, we have covered the main conditions determining the development of CRs resulting from the CS signaling the occurrence of the US. However, subjects can also learn that a CS signals the *omission* of a US. When the CS signals US delivery, we talk about *excitatory* classical conditioning or *conditioned excitation* (presentation of the CS "excites" the production of a CR). In contrast, when the CS signals US omission, we talk about *inhibitory* classical conditioning or *conditioned inhibition* (presentation of the CS "inhibits" the production of a CR). This section briefly describes the procedures used to produce and measure conditioned inhibition.

Procedures That Yield Conditioned Inhibition

Pavlov's Procedure

In this procedure, first described by Pavlov (1927), the researcher pairs a CS with the US to create a conditioned excitor (i.e., a CS that produces a CR). These trials take the form of CS_{exc}-US. In other trials, a second CS (CS_{inh}) is presented together with CS_{exc} and the pair is not followed by the US. These trials take the form of $CS_{exc}CS_{inh}$-noUS. Thus, subjects learn that CS_{exc} predicts the US and they produce CRs when CS_{exc} is presented. But they also learn that whenever CS_{inh} is present, no US will be delivered, even if CS_{exc} is also present. Thus, the CR that should be produced by CS_{exc} is not observed because of the presence of CS_{inh}. That is, the presence of CS_{inh} attenuates the CR that would otherwise be elicited by CS_{exc} (see Figure 33.4). This procedure is commonly used to produce conditioned inhibition and, for that reason, it is sometimes referred to as the *standard procedure*.

All of us have been subject to the Pavlov procedure for producing conditioned inhibitors in our daily lives. For example, when I call my dog and say "Dinner!" (CS_{exc}) he soon gets a meal (US), and this food makes him salivate profusely (UR). With repetition, he has come to salivate profusely (CR) whenever he hears the word "Dinner!" However, he knows that if I say "Dinner!" when we are at the dog park (CS_{inh}) he does not salivate because he has come to realize that a meal is at least one hour away (i.e., no US follows). Thus, the situation involves CS_{exc}-US ("Dinner!" signaling food) trials and $CS_{exc}CS_{inh}$-noUS ("Dinner!" + dog park signaling no food) trials. As a result of this training, CS_{exc} becomes excitatory (it produces the salivation CR) and CS_{inh} becomes inhibitory (it does not produce the CR).

Other Procedures

There are other types of training that may result in conditioned inhibition. The inhibition resulting from these procedures is not as robust as that resulting from the standard and negative contingency procedures; thus, they are not that widely used. However, it is worth mentioning that the use of a *negative contingency* (presenting the CS and US more often apart than together; Rescorla, 1968) can also result in the development of inhibition because subjects learn that the presence of the CS signals the absence of the US. Similarly, *backward conditioning* can result in the development of conditioned inhibition to the CS. Although a few US-CS pairings lead subjects to produce a CR whenever the CS is presented (subjects act as if they know the US is expected some time around the CS), many US-CS pairings lead subjects to realize that the CS does not signal US delivery but rather US termination (Heth, 1976). Thus, subjects come to expect no US when the CS is presented, and inhibition develops.

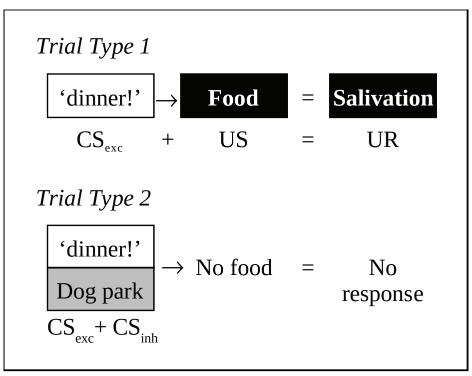

Figure 33.4 Pavlov's (standard) procedure for producing conditioned inhibition. As a result of receiving both trial types during training, CS_{inh} comes signal US omission and consequently inhibits the CR.

Measurement of Conditioned Inhibition

The Measurement Problem

Assessing the development of conditioned excitation is relatively simple: All we have to do is watch for a CR. However, measuring conditioned inhibition is not that simple. We need to find a way to measure learning of *inhibition* (i.e., nonproduction) of a response. How can we measure something that does not occur? Researchers have come up with several strategies, the most common of which are known as summation and retardation tests.

Summation Tests for Conditioned Inhibition

Conditioned inhibitors inform subjects that an expected US will not be delivered. Thus, they should attenuate expectation of US delivery when presented together with an excitor. That is, if we present subjects with an excitatory CS_{exc} alone, they should exhibit a CR of a certain magnitude. However, if we present subjects simultaneously with excitatory CS_{exc} and inhibitory CS_{inh} (i.e., $CS_{exc}CS_{inh}$), the magnitude of CR that we observe to CS_{exc} should be dramatically reduced. Thus, if my saying the word "Dinner!" (CS_{exc}) makes my dog salivate (CR), but saying the word "Dinner!" at the dog park (CS_{inh}) does not, we can begin to assume that the dog park acts as an inhibitor of the salivation CR. Importantly, the CS_{exc} that we use for summation testing can be different from the one used during training. To test this assumption, I can use a different word that also produces the salivation response such as "Breakfast!" ($CS_{exc'}$). If using this word in the dog park results in little salivation, we can conclude that our CS_{inh}, the dog park, is indeed a conditioned inhibitor. You can conceptualize summation as a situation in which CS_{exc} needs to counter the inhibitory power of CS_{inh} before a CR can actually be observed.

Retardation Tests for Conditioned Inhibition

The term *retardation* refers to speed of acquisition of a CR. It is assumed that, following conditioned inhibition training, an inhibitory stimulus will require more pairings with a US to become a reliable excitor than a stimulus that is neutral at the start of the CS-US pairings. Thus, if I decide to feed my dog (US) at the dog park (CS_{inh}) and at the small field behind my house ($CS_{neutral}$), it should take longer for him to salivate when getting to the dog park (a CS that signals US omission) than when getting to the small field (a novel stimulus) behind the house. Salivating when getting to the small field reflects the acquisition of a CS-US association in the absence of inhibition; the delay in salivating when getting to the dog park reflects retarded acquisition. You can conceptualize retardation as a situation in which CS_{inh} needs to overcome inhibition before it can begin acquiring excitation.

So, Which Test Is Definitive?

Neither. Factors other than conditioned inhibition could explain what is observed following summation and retardation tests. Fortunately, the alternative explanations for summation and retardation are mutually exclusive. (See Rescorla, 1969, for elaboration.) Thus, researchers have resorted to something known as the *two-test strategy*: To consider a CS_{inh} as a true inhibitor, it must "pass" (i.e., prove to be inhibitory with) both a summation test and a retardation test.

HIGHER-ORDER CLASSICAL CONDITIONING

So far, we have talked about conditioning situations in which the US is a stimulus that elicits an innate response (i.e., a reflex) such as food producing salivation or spoiled food producing stomach illness. But that is not the end of all conditioning. Indeed, most of our conditioning experiences involve learning about stimuli that do not produce innate, reflexive responses. Let us explain this with an example. Mary (CS) broke up (US) with Steven in the fall, which made him sad (UR). Whenever Steven sees Mary (CS), he feels angry (CR). In the spring, Steven and Mary are taking a class together. He sees Mary talking to Julia every day. Now he also feels angry whenever he sees Julia. Notice that in this example, Julia's presence has not been paired with any unpleasant event that should result in elicitation of the anger CR. However, she has become associated to Mary, a CS that already elicited the CR (see Figure 33.5).

Higher-order conditioning occurs whenever a stimulus becomes a CS because it has been paired with another CS that already elicits a CR. In our example above, Mary is a *first-order CS* because it has a first-degree association to the US (it is only one step removed from the US). In contrast, Julia is a *second-order CS* because it has a second-degree association to the US (it is two steps removed from the US). Third, fourth, and other higher orders can be conditioned by pairing a new CS with the previous order CS (e.g., pairing Julia with Stella would make Stella a third-order CS).

Do CS-US Pairings Always Result in CRs?

In a word, the answer to this question is *no*. Sometimes, you will not observe a CR even if the CS and US are repeatedly paired. Most of these situations can be grouped into situations in which the CS or US has become too familiar and situations in which another stimulus has become a better or more reliable predictor of the US.

Situations in Which the CS or US Have Become Too Familiar

It seems counterintuitive, but if a subject is too familiar with the CS or US, conditioning is not as effective as if the CS and US were new. For example, cancer patients tend to develop aversions (CRs) to foods (CSs) consumed around the time of their chemo or radiotherapy treatments (US) because these treatments make them feel very nauseous. Interestingly, they seem to take longer to develop aversions to highly familiar foods, like bread, than to less familiar foods. Situations like this, in which the CS is repeatedly presented alone before the CS-US pairings, reflect the so-called *CS-preexposure effect* (also known as *latent inhibition*; Lubow & Moore, 1959). In other situations, it is not the CS but the US that is too familiar. Thus, a cancer patient undergoing chemotherapy treatment is frequently exposed to chemicals (US) that result in nausea (UR). Later in treatment, this patient may have a hard time realizing that nausea is related only to receiving a red chemical and not a yellow chemical (CSs) because in her experience the nausea always occurs, even if nothing predicts it. This situation reflects the so-called *US-preexposure effect* (Randich & LoLordo, 1979).

Thus, if either the CS or the US is repeatedly presented alone before the CS-US pairings take place, acquisition of the CS-US association will exhibit *retardation* (i.e., delay in producing CRs when presented with the CS). Note that this retardation is functionally equivalent to the retardation that follows conditioned inhibition (see the *Measurement of Conditioned Inhibition* section); however, conditioned inhibition is not assumed to develop in the preexposure situations because preexposed CSs do not "pass" summation tests for conditioned inhibition (i.e., they do not conform to the two-test strategy; Rescorla, 1971).

Situations in Which Another Stimulus Is a Better or More Reliable Predictor of the US

In all learning situations we encounter more than one stimulus that can become associated to the US. In most of these situations, however, the

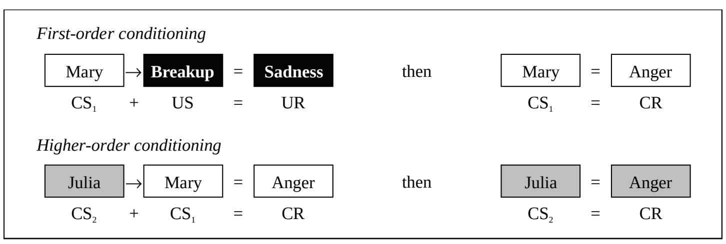

Figure 33.5 Higher-order conditioning.

NOTES: CS_1 (Mary) represents the first-order CS, which is paired with the US and consequently produces a CR. CS_2 (Julia) represents the second-order CS, which is paired with CS_1 and consequently produces a CR.

great majority of the stimuli are somewhat irrelevant to the situation, are very familiar to the subject, or go unnoticed by the subject. Among those stimuli that the subject notices, there are differences in the degree of conditioning they acquire. Theorists usually say that these stimuli *compete* for association with the US and call the situations described below *stimulus competition*.

In a Halloween haunted house, a young child is scared by a screeching bat (CS_{high}) and a quiet spider (CS_{low}) that suddenly jump out of a pantry (US). Later on, he seems scared of bats but not spiders. Potentially, the characteristics of the CSs made a difference in this learning experience: One of the two CSs, the bat, was more salient than the other, the spider. Usually, when one of the CSs that predict the US is more salient than the others, the salient CS tends to acquire CR-eliciting properties that the other CSs do not acquire. This phenomenon is known as *overshadowing* (Pavlov, 1927). The CS_{high} is known as the *overshadowing CS*, whereas the CS_{low} is known as the *overshadowed CS*. Why overshadowing occurs is still under debate. Some researchers say that subjects simply attend to CS_{high} and ignore CS_{low}, whereas others say that when subjects have to produce a CR they look at the learning experience and select the best predictor of the US (due to salience, this would be CS_{high}). Whatever the theoretical explanation, overshadowing is a reliable phenomenon observed in multiple situations.

Imagine that the same child went to a haunted house the previous year and was terrified of a spider (CS_{old}) that fell off the ceiling onto his head. Now, when exposed to the spider (CS_{old}) and the bat (CS_{new}), the kid fears the spider and not the bat. In this case, due to prior experience, the spider (CS_{old}) has become a better predictor of a fear-producing US. The added stimulus, the bat, is then disregarded as a predictor of the US because the already established predictor, the spider, is present. In these types of situations, we say that the previous experience with the spider (CS_{old}) predicting the US resulted in *blocking* of learning that the bat (CS_{new}) also predicted the US (Kamin, 1968). Just as in the case of overshadowing, there are multiple theoretical explanations of blocking; despite their differences, all of them are based on the idea that CS_{old} is viewed by subjects as a more reliable predictor of the US than CS_{new}.

BIOLOGY VERSUS EXPERIENCE: SOME CONSTRAINTS ON LEARNING

Pavlov suggested that classical conditioning could be somewhat arbitrary. Indeed, he suggested that any stimulus could become associated to any US if they were paired. However, that is not always the case. Subjects' biology takes a primary role when determining whether two stimuli will become associated. These *biological constraints on learning* are viewed as a result of the evolutionary history of the organism, and they reflect the most advantageous way to adapt to their ecological niche. Thus, some situations benefit from the *belongingness* between the CS and US (biologically, they are meant to be associated), and few or even one CS-US pairing result in strong CRs. This is also known as *preparedness* (animals' readiness to acquire some associations faster than others).

The issue of belongingness was beautifully investigated by John Garcia and Robert Koelling in 1966. In their preparation, rats were presented with a drinking tube that delivered flavored water. Additionally, every time the animal extended its tongue to drink from the tube, a light and sound were turned on. Thus, the experience of the animals was of "bright-noisy-flavored water"; this constituted CS_{bnf}. Then, half of the animals were exposed to radiation, which made them ill to their stomach ($US_{radiation}$). The other half received electrical shock immediately after drinking from the tube (US_{shock}). Then, each of these two groups was again divided in half. Some of the animals received the "bright-noisy" component of the CS (CS_{bn}; plain water accompanied by the light and sound), whereas the remaining animals received the flavor component of the CS (CS_f; flavored water with no light or sound). The procedure of the experiment is summarized in Figure 33.6.

The question was which subjects would refuse to drink from the tube? To anticipate the answer to this question, you must think about the world of a rat, which is composed mostly of dark burrows. If you were a rat, how would you find food in your burrow? Probably through smell. And how would you avoid danger? Probably by carefully listening to sounds and watching out for drastic changes in luminosity of your environment. After considering this, the results of the study will not seem so surprising: Rats that received CS_{bnf}-$US_{radiation}$ pairings refused to drink from

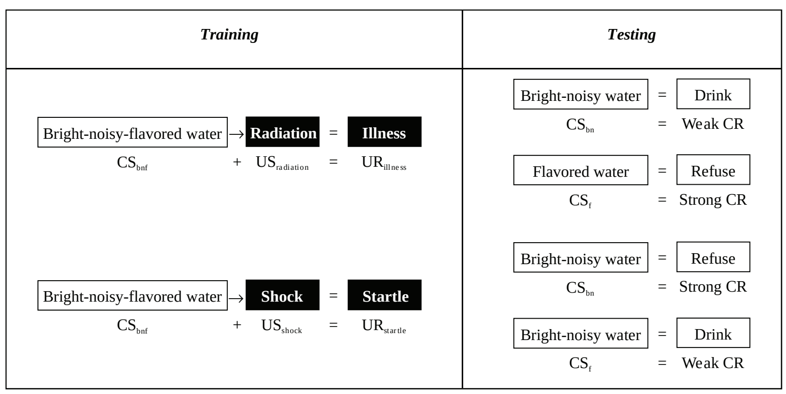

Figure 33.6 Schematic representation of Garcia and Koelling's (1966) study.

the spout only when presented with flavored water (CS_f). In contrast, rats that received CS_{bnf}-US_{shock} pairings refused to drink from the spout only when presented with the light and sound cues (CS_{bn}). Thus, when it came to avoiding illness, they used the cues that they would normally use in their environment to determine whether something is edible—namely, smell and taste. However, when it came to avoiding danger, they used as cues sound and changes in illumination. These basic observations apply across species and, as we will see in the final section of this chapter, they are prevalent in many daily situations.

CLASSICAL CONDITIONING IN ACTION: DEVELOPMENT AND TREATMENT OF FEARS AND PHOBIAS

In 1920, John B. Watson and his assistant Rosalie Rayner provided the first empirical evidence that fears can be acquired through classical conditioning. Their subject, 11-month-old Albert, was allowed to play with a white rat (CS). While Albert was playing with the rat, a loud noise (US) occurred in the background. Albert became startled and cried when the loud noise occurred in the background (UR). After several CS-US pairings, Albert was presented with the rat and no noise occurred in the background (this was done to assess the strength of the CR). Albert cried and tried to get away from the rat (fear CR). Furthermore, this fear response was observed also when presented with other stimuli that somehow resembled the rat: He was fearful of other animals (dog, monkey, and rabbit) and of white "fluffy" things (Santa Claus mask). That is, the fear CR *generalized* to other, similar stimuli.

The "little Albert" study exemplifies the origin of most fears and phobias: an event that initially did not produce fear can become fear-producing when paired with another event that unconditionally produces a fear response. Note that most people who experience phobias do not remember the conditioning episode that got them started. However, that does not undermine the fact that classical conditioning may be at the root of that particular phobia.

Observational Learning and the Development of Fear

The principle of belongingness suggests that not all CS-US pairings will result in the development of phobias. We are better prepared to fear some things (e.g., spiders, snakes, and heights) than others (e.g., fruits, rocks, and flowers), probably because our evolutionary history made us more attuned to danger in some situations. (Note that with an intense enough experience, some people may develop fear of fruits, rocks, and flowers.)

An extraordinary example of this principle was provided by Susan Mineka and her colleagues (e.g., Mineka & Ben Hamida, 1998). They presented a wild-reared rhesus monkey (the model) with two objects, a toy snake and a toy flower, and videotaped his reactions. This monkey had experience with both snakes and flowers and exhibited strong fear reactions to the snake but not to the flower. Then, they presented lab-reared monkeys (the observers), which had no experience with snakes, with the videotapes. For some observers, the tape was of the actual model-snake and model-flower encounters. Then, the observers were presented with the toy snake and the toy flower and, not surprisingly, they exhibited fear of the snake but not the flower. That is, *observational learning* of the fear reaction had taken place. Interestingly, other observers were presented with edited tapes in which the snake and flower had been digitally switched (thus, the model exhibited fear of the flower but not the snake). When presented with the toy flower and the toy snake, these observers displayed fear of neither object. That is, monkeys (like humans) are predisposed to fear snakes (but not flowers) and a single observational experience with a snake (but not a flower) is enough to trigger fear responses. It is believed that observational learning is at the root of many common fears.

Classical Conditioning and the Treatment of Fears and Phobias

So far, we have established that CS-US pairings may be the cause of common fears and phobias. How can we get rid of a CR established through CS-US pairings? The most common procedure is *extinction* (presenting the CS without the US to attenuate the CR). In therapeutic settings, this procedure is known as *exposure therapy*, and it basically consists of presenting the subject with the feared object repeatedly to gradually attenuate the fear response. Although there are some techniques that expose subjects to the CS for long periods of time without allowing them to escape the situation (known as *flooding*; Baum, 1970), most therapists expose subjects to the CS gradually, starting with asking the subject to imagine the object and gradually presenting more realistic versions of the object (e.g., pictures, models, etc.) until the subject manipulates the actual feared object. Another commonly used technique involves conditioning an opposite CR to the object, a procedure known as *systematic desensitization* (Wolpe, 1969), which is based on the classical conditioning principle of *counterconditioning*. For example, if a subject fears a dog, a picture of a dog will be used as a CS paired with a US of relaxation. Because relaxation and fear are incompatible responses, the relaxation response will eventually overcome and replace (i.e., counter) the fear response.

A major problem with these techniques is that *relapse* (the return of fear) is very prevalent. The processes of *renewal* and *spontaneous recovery* described previously are some of the major sources of relapse. Current research is trying to identify the major variables determining the return of extinguished responses, as well as the best strategies to prevent relapse.

SUMMARY

Pavlov's studies demonstrated that even basic, reflexive responses can come under control of experience. Indeed, classical conditioning is prevalent in many aspects of daily life, including the development of emotional responses, food preferences and aversions, and development of fears and phobias. Conditioned responses (CRs) can be produced through CS-US pairings, eliminated by exposing subjects to the CS without the US (CS-noUS trials), or inhibited in situations in which a CS_{inh} signals the omission of an otherwise expected US. Importantly, CRs can be transferred to other, novel CSs just by pairing them with a CS previously paired with a US; thus, learning can occur even in the absence of USs. Conditioning is most effective when the CS and US are relatively novel, occur in close proximity, occur reliably with each other, and occur intensely enough to prevent overshadowing by other stimuli. But experience is limited by biology, and some associations are easier to form than others. These principles allow us to better understand which stimuli in the environment are more likely to control the behavior of organisms, and help us devise better strategies for acquisition and elimination of behaviors through classical conditioning.

REFERENCES AND FURTHER READINGS

Baum, M. (1970). Extinction of avoidance responding through response prevention (flooding). *Psychological Bulletin, 74*, 276–284.

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10*, 445–466.

Domjan, M. (2003). *The principles of learning and behavior.* Belmont, CA: Wadsworth/Thomson.

Escobar, M., & Miller, R. R. (2004). A review of the empirical laws of basic learning in Pavlovian conditioning. *International Journal of Comparative Psychology, 17*, 279–303.

Estes, W. K., & Skinner, B. F. (1941). Some quantitative properties of anxiety. *Journal of Experimental Psychology, 29*, 390–400.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4*, 123–124.

Gunther, L. M., Miller, R. R., & Matute, H. (1995). CSs and USs: What's the difference? *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 15–30.

Hearst, E. (1975). Pavlovian conditioning and directed movements. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 215–262). New York: Academic Press.

Heth, C. D. (1976). Simultaneous and backward fear conditioning as a function of number of CS-UCS pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 117–129.

Hollis, K. L. (1997). Contemporary research on Pavlovian conditioning: A "new" functional analysis. *American Psychologist*, *52*, 956–965.

Kamin, L. J. (1965). Temporal and intensity characteristics of the conditioned stimulus. In W. F. Prokasy (Ed.), *Classical conditioning* (pp. 118–147). New York: Appleton-Century-Crofts.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior: Aversive stimulation* (pp. 9–31). Miami, FL: University of Miami Press.

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced preexposure to the conditioned stimulus. *Journal of Comparative and Physiological Psychology, 52*, 415–419.

Matzel, L. D., Held, F. P., & Miller, R. R. (1988). Information and expression of simultaneous and backward associations: Implications for contiguity theory. *Learning and Motivation, 19*, 317–344.

Mineka, S., & Ben Hamida, S. (1998). Observational and nonconscious learning. In W. T. O'Donohue (Ed.), *Learning and behavior therapy* (pp. 421–429). Needham Heights, MA: Allyn & Bacon.

O'Donohue, W. T. (1998). *Learning and behavior therapy*. Needham Heights, MA: Allyn & Bacon.

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.

Randich, A., & LoLordo, V. M. (1979). Preconditioning exposure to the unconditioned stimulus affects the acquisition of a conditioned emotional response. *Learning and Motivation, 10*, 245–277.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66*, 1–5.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin, 72*, 77–94.

Rescorla, R. A. (1971). Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology, 75*, 77–81.

Rescorla, R. A., & Cunningham, C. L. (1979). Spatial contiguity facilitates Pavlovian second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *5*, 152–161.

Sclafani, A. (1997). Learned controls of ingestive behaviour. *Appetite, 29*, 153–158.

Shanks, D. R. (1994). Human associative learning. In N. J. Mackintosh (Ed.), *Animal learning and cognition* (pp. 335–374). London: Academic Press.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology, 3*, 1–4.

Wolpe, J. (1969). *The practice of behavior therapy*. New York: Pergamon Press.

34

RECENT TRENDS IN CLASSICAL CONDITIONING

JAMES C. DENNISTON
Appalachian State University

When most people hear the words "Pavlovian conditioning" or "Pavlov" they conjure a mental image of a dog salivating to the sound of a bell that has been previously paired with food. Although such an awareness of this basic learning process is surprising widespread, it is equally noteworthy that most people's understanding of conditioning is restricted to this particular example. Pavlovian conditioning leads to a rich form of learning that allows animals to anticipate, and consequently prepare for, important events in their environment. Over the past few decades, our knowledge of the nature and mechanisms of Pavlovian conditioning has evolved dramatically from the layperson's understanding of the conditioning process to that of a rich form of associative learning.

LAWS OF LEARNING

A dominant early view of the conditioning process was that a handful of general principles, or *laws of learning*, were capable of explaining the conditioning process. Many of these laws of association were first suggested by philosophers, including Aristotle and David Hume (1739/1964). One such law is the *law of contiguity*, which states that an association will be formed between stimuli when these stimuli occur close together in time. Applied to Pavlovian learning, pairings of a conditioned stimulus (CS; e.g., a bell) with an unconditioned stimulus (US; e.g., food) generate the strongest conditioned responding to the CS when researchers present these stimuli close together in time (a procedure known as *short-delay conditioning*). However, equally important is the *law of contingency*, which states that the occurrence of the effect is dependent upon the occurrence of the cause—in other words, the US should only occur along with the CS. Any weakening of the CS-US contingency has a deleterious effect on conditioned responding. Take, for example, my dog learning to bark at the sound of my doorbell ringing. The first few times she heard the doorbell ring, there was no response. However, after several episodes of ringing being followed closely by my greeting people at the door, she began to bark whenever she heard a doorbell ring. In this particular example, that the doorbell was closely followed by people entering our home demonstrates temporal contiguity, whereas the consistency of the doorbell being followed by people entering our home (and entering our home only when the doorbell was rung) reflects the strong contingency between these events. Sometimes, however, my child watches a particular cartoon on television in which a doorbell rings with some frequency. At first, my dog would respond strongly (and loudly) to the sound of this doorbell, just as it did with ours. However, presentation of the doorbell in the absence of visitors weakens the CS-US contingency, in that the CS is sometimes experienced without the US. Likewise, some visitors merely walk into our home without ringing the bell, thereby leading to presentation of the US in the absence of the CS. Both of these examples reduce the

CS-US contingency and consequently the likelihood of my dog barking at the sound of doorbells, because it is no longer as reliable a predictor of people entering the house as it had been. Researchers have long believed that these two factors, contiguity and contingency, are two of the major determinants of Pavlovian learning.

THE FUNCTION OF PAVLOVIAN CONDITIONING

Associations formed through Pavlovian conditioning can provide significant adaptive value to organisms in our environment. For example, learning to avoid foods that have previously made us sick (taste-aversion learning), identifying the sounds of an approaching predator, or developing a fear of potentially harmful animals (e.g., snakes or spiders) are the consequence of Pavlovian conditioning. Being able to form such associations provides animals with an adaptive edge in that avoidance of potentially dangerous stimuli will increase the likelihood that an organism will successfully reproduce and therefore pass on its genes to future generations. When one considers the value of Pavlovian conditioning from a functional perspective, it is clear that this form of learning extends far beyond learning to salivate to bells. Instead, Pavlovian learning allows animals to anticipate important events in their environment (e.g., food, water, sex, painful stimuli, and illness). Once an animal has learned to anticipate such events, this knowledge then allows the animal to prepare for these events though a variety of behaviors known as *conditioned responses*. As a result of this process, Pavlovian learning is responsible for the emotional responses that we have to many of the stimuli in our environment, whether it be life-or-death anticipatory responses to stimuli that signal impending danger, or the (hopefully) happy feeling that overcomes you when you hear the song that you first danced to with your significant other. (Notably, if your relationship with your significant other sours, so too will your response to the song, a process known as *counterconditioning*, in which changes in the value of a US result in corresponding changes in the value of the CS.)

HIGHER-ORDER CONDITIONING

Second-Order Conditioning

Interestingly, Pavlovian learning is not restricted to biologically significant events such as food and painful stimuli. Instead, the mechanisms underlying Pavlovian conditioning allow animals to associate nearly all events that occur close together in time. Take, for example, a group of phenomena known as *higher-order conditioning*. One example of higher-order conditioning is *second-order conditioning* (Rizley & Rescorla, 1972), in which a previously conditioned stimulus functions much like a US. Procedurally, second-order conditioning occurs when a CS, which we will call CS1, is conditioned through pairings with the US (i.e., CS1→US, a process often referred to as *first-order conditioning*) and then a second CS, CS2, is paired with CS1 (i.e., CS2→CS1). Importantly, these latter CS2-CS1 pairings occur without the US present. As a consequence of this training, CS2 is capable of producing a conditioned response. Second-order conditioning is interesting for a couple of reasons. First, the phenomenon demonstrates that conditioning it not restricted to CS-US pairings, as CS2 is conditioned as a result of its pairings with another (previously conditioned) CS. As the name of this phenomenon implies, CS1 has a first-order (direct) association with the US, whereas CS has a second-order relation with the US (through CS1). Second-order conditioning is particularly interesting from an applied perspective, as it allows one to develop fears of stimuli that were never directly paired with a US. For example, imagine that you are afraid of snakes (a previously conditioned CS) and one day you take a hike and come across a snake on a particular trail. When you encounter the snake you become somewhat fearful and quickly retreat from the snake without anything bad happening to you (this is effectively a pairing of CS2, the trail, with CS1, the snake). The next time that you hike on this particular trail, you are likely to experience anxiety when you approach the location where you previously observed the snake. This process of higher-order conditioning allows conditioned fears to spread to stimuli as they become associated with a previously fearful stimulus, which is a potential reason for why my wife no longer takes out the garbage following an experience in which she found two mice (which she does not care for) in our garbage can!

Sensory Preconditioning

Another example of higher-order conditioning is *sensory preconditioning* (Brogden, 1939). Sensory preconditioning is similar to second-order conditioning, except that the order of learning is reversed, such that two neutral stimuli, CS2 and CS1, are paired together and then one of the stimuli, CS1, is made biologically important through pairings with the US. As a result of this training, CS2 becomes capable of producing a conditioned response. The primary difference between sensory preconditioning and second-order conditioning is the status of CS1 at the time of CS2-CS1 pairings. In sensory preconditioning, CS1 is neutral, whereas in second-order conditioning, CS1 is already conditioned. This procedural difference demonstrates that in the case of sensory preconditioning, learning can occur between two neutral stimuli. In other words, the presence of a US is not necessary for learning to occur. Instead, the US merely provides a reason for the animal to demonstrate what it has learned, a relation between CS2 and CS1 and between CS1 and the US. When most people think about the learning process, they commonly assume that we are most likely to learn about important stimuli and not seemingly trivial events in our environment. Sensory preconditioning demonstrates that Pavlovian learning allows us to associate seemingly unimportant events in our environment. Take, for example, a child who frequently goes to a park where many people walk their dogs. The presence of dogs at the park is likely to be associated. Should the child one day be attacked by a dog in a different place, the child is now likely to demonstrate a reluctance to go to the park that was previously associated with dogs. Examples of this nature point to the fact that what is trivial now might well be critically important at some time in the future. Were learning to be restricted to biologically important events, we would lack many of the potentially important predictors of events in our environment.

TIMING OF CONDITIONED RESPONDING

Pairings of a CS and a US are thought to result in the formation of an association between the two stimuli. This Pavlovian association is responsible for the ability of the CS to produce a conditioned response. Since Pavlov's (1927) pioneering work, researchers have known that the conditioned responding can be controlled by a variety of factors, including the temporal (time) relation between the CS and the US. For instance, Pavlov found that following pairings of a long-duration CS with a US, conditioned responding was not constant throughout the CS. Instead, very little responding was observed during the early portions of the CS, but at the time at which the US was normally presented, conditioned responding increased. Pavlov termed this effect "inhibition of delay," as he believed that the early portions of the CS signaled that the US was still some time away and consequently "inhibited" or prevented the conditioned response, whereas the latter portions of the CS were "excitatory" and triggered the conditioned response. Such effects demonstrate that the temporal relation is part of the knowledge gained through Pavlovian conditioning and that this information can control when an animal responds.

Researchers have demonstrated temporal control of conditioned responding in a number of different studies. For example, Rosas and Alonso (1997) presented rats with a long duration CS (a 150-s tone) followed by a brief, mild footshock to condition a fear of the CS. Early in training, subjects demonstrated a fear of the tone that was constant throughout the entire CS. However, as training progressed, very little fear was observed during the early portions of the CS, whereas greater fear was observed toward the end of the CS, a result that replicates Pavlov's inhibition of delay experiments. Interestingly, when researchers tested subjects after various delays between training and testing, the temporal control of conditioned responding was lost. That is, a few weeks after training, fear was strong throughout the CS. This finding suggests that the rats remembered the CS-US pairing, but gradually forgot the temporal relation. More generally, temporal information that influences when an animal responds appears to be independent from the more general association that influences whether an animal will respond to a CS.

INHIBITORY LEARNING

Pavlov's (1927) pioneering work on the study of associative learning focused not only on animals learning to anticipate the occurrence of events in their environment (what Pavlov called *excitatory conditioning*), but also on the ability to learn that some stimuli predict the absence of an event. Learning about the omission of a US was termed *conditioned inhibition*, as this form of learning led to the inhibition of a behavior that would have otherwise occurred in the presence of a CS. In a typical conditioned inhibition procedure, a CS is paired with the US except when another stimulus is present. This second stimulus becomes a conditioned inhibitor, as it is thought to signal the absence of the otherwise expected US.

Barnet and Miller (1996) investigated the content of inhibitory learning using rats as subjects. In their experiment, two types of conditioning trials were presented. In one type of trial, the excitatory CS was paired with a brief footshock in order to condition a fear of the CS. On other trials, the excitatory CS was presented along with the inhibitory CS and the shock US was omitted. Barnet and Miller expected that the excitatory CS would come to signal the presentation of the US at a particular moment in time (based upon studies like those described previously). Their question was whether the inhibitory CS would come to signal the omission of the US at a particular moment in time. It is worth noting that this is a particularly difficult question to study because it is relatively easy to examine the timing of a fear response, as the animal's fear response can be measured throughout the presence of the CS, whereas the ability to measure an absence of fear is complicated by the fact that the inhibition of fear results in no explicit change in behavior. That is, when an inhibitory CS is presented by itself (and the animal presumably expects that the shock will not occur), the animal typically does not change its behavior in any noticeable way. Simply put, if you do not expect something bad to happen, you merely continue on with what you were doing. In order to deal with this measurement issue, Barnet and Miller used what is called a *summation test for conditioned inhibition* (Rescorla, 1969). In a summation test, the potential of the inhibitor to reduce an animal's fear response to another CS is assessed. If the supposed inhibitor truly signals that the US will not occur, then it should be able to reduce fear triggered by some other fearful stimulus (i.e., it transfers its inhibitory properties to another stimulus). In their experiment, Barnet and Miller presented their subjects with two types of test trials. In the first, the inhibitor was presented along with a test stimulus that signaled the presentation of the US at the same time at which the subjects were thought to have developed an expectation for the absence of the

US (based upon the training already described). In the second, subjects were presented with the inhibitor along with a test stimulus that signaled the presentation of the US at a different time than that at which the subjects were thought to have developed an expectation for the absence of the US. Their results revealed that fear was inhibited only when these two expectations (presentation of shock and omission of shock) occurred at the same time. These findings suggest that an inhibitor not only tells the animal that the US will not occur, but also signals when the US will not occur—thereby resulting in a stimulus that signals safety for a limited time period.

EXTINCTION OF CONDITIONED RESPONDING

Pairings of a CS and a US result in the acquisition of a conditioned response, a process that alters the way in which we respond to the CS. Likewise, presentations of the CS in the absence of the US results in a decrease in conditioned responding to the CS, a phenomenon termed *extinction*. Extinction of Pavlovian responding can be analyzed in three ways—operationally, empirically, and theoretically. Operationally, extinction simply involves the presentation of the CS without the US. Empirically, extinction results in a decrease in conditioned responding to the CS. In contrast to these two clear-cut means of understanding extinction, less agreement is reached at the theoretical level of analysis of why the conditioned response decreases when the CS is presented alone. Superficially, one might claim that the response deceases simply because the US is no longer presented, but this merely begs the question of why the response diminishes (or more precisely, what is the nature of learning that occurs during extinction?). Such a question has both great theoretical and implied implications given the pervasive use of extinction procedures to treat anxiety disorders.

Theories of Extinction

One simple theoretical explanation of extinction is that presentations of the CS in the absence of the US lead to a weakening of the CS-US association. This explanation essentially claims that the association is unlearned, effectively returning the CS to a neutral status (as if it had never been previously conditioned). However, several findings are problematic for this explanation. First, Pavlov (1927) found that following successful extinction of a CS, conditioned responding returns when a delay is imposed between extinction and subsequent testing, a phenomenon called *spontaneous recovery* (as the response recovers on its own following a delay). That the conditioned response is capable of returning demonstrates that the original excitatory association between the CS and US must still be intact, thereby requiring an alternative explanation for the loss of responding caused by the extinction procedure. A second phenomenon that is often observed is that when a previously extinguished CS is reconditioned through additional pairings of the CS with the US, the conditioned response emerges faster than during initial acquisition training, a phenomenon termed *facilitated reacquisition* (Konorski & Szwejkowska, 1950, 1952). Again, this typically more rapid reacquisition of conditioned responding implies that the original CS-US association must still be intact. Finally, environmental changes can alter the effectiveness of extinction, a phenomenon termed the *renewal effect*.

Renewal of Conditioned Responding

Bouton and other researchers have spent many years studying the effect of environmental cues on extinction of conditioned responding. Bouton and Bolles (1979) provided rats with fear conditioning through pairings of a CS with a brief shock in a particular type of experimental chamber (or Skinner box). Following acquisition training, subjects received extinction of the CS in a different type of Skinner box. Typically, the experimental chambers are altered such that they appear, smell, and feel different through the addition of textured floors, patterned walls, odors, changes in illumination, and often differently shaped or sized chambers in order to alter as many sensory characteristics of the training environment (or context) as possible. Following extinction of the CS, subjects were returned to the original training environment in order to measure their fear of the CS. This procedure is known as an *ABA renewal procedure,* as acquisition training is provided in Context A, extinction in Context B, and testing in Context A. At test, subjects demonstrated a return (or renewal) of fear to the CS. This effect is somewhat different from that of spontaneous recovery, as the return of conditioned responding occurs as a consequence of a change in physical context rather than a delay imposed between extinction and testing. That the conditioned response returns when testing occurs outside of the place in which extinction was provided once again demonstrates that the original excitatory association is somehow preserved during extinction.

Why extinction is disrupted with a change of context has been a hot topic of research over the past several decades. One potential explanation for the renewal effect is that when subjects are tested for renewal of conditioned responding in the context in which acquisition training was provided (Context A), that context “reminds” the animal of the fear conditioning that occurred in that setting, thereby producing a return of fear. However, this appears to be only a partial explanation for this effect. In another experiment, Bouton and Bolles (1979) found that when fear conditioning was provided in Context A followed by extinction in Context B, renewal of responding was also observed when testing was conducted in a new context, Context C (an *ABC renewal design*). This finding demonstrates that renewal is likely to be observed whenever testing with the CS occurs outside of the extinction context, suggesting that animals fail to remember the extinction experience once they leave the place in which extinction was provided. Notably, ABC renewal is typically somewhat weaker than ABA renewal, providing some evidence that a return to the fear-conditioning context does help remind the animal of the original fear. This reminder, in addition to a failure to fully generalize extinction, leads to a somewhat stronger return of fear.

At the theoretical level, researchers have offered several explanations of the renewal effect. One possibility is that inhibitory learning (which is thought to occur during extinction) does not generalize as well as excitatory learning (i.e., the original fear). Such an explanation is attractive from an evolutionary perspective as it might be safer for one to assume that something harmful will occur than to assume that the potentially harmful event will not occur. Despite the attractiveness of this theoretical perspective, recent evidence casts doubt upon this view. Nelson (2002) pointed out that in a typical renewal procedure, extinction, and consequently the formation of an inhibitory association, is always the second experience that an animal learns. Hence, inhibitory learning and order of training are confounded, such that it is difficult to separate whether it is inhibitory learning rather than second learned information that becomes context specific. To investigate this question, Nelson reversed the typical order of training so that rats received inhibitory training first and excitatory training second. When they tested subjects in a different context, the researchers found that the first learned association generalized and that the second learned association did not. Hence, second learned information (whether it be excitation or inhibition) tends to be more context specific than first learned information.

Toward addressing why second learned associations become context specific, Bouton (1993, 1994) claimed that extinction of a CS results in multiple memories—one consisting of the original excitatory association and another consisting of an inhibitory CS-US association. Hence, the CS becomes ambiguous, as it now has two meanings. In other words, sometimes the CS is followed by the US and other times it is not. This effect is much like words in the English language that can have multiple meanings—for example, the word "fire" can mean very different things depending upon whether you are in a burning building or standing in front of a firing squad (Bouton, 1997). When we encounter ambiguous words, we use surrounding (context) words to clarify meaning. Likewise, animals use their physical surroundings to determine the current meaning of a CS following acquisition and extinction training. According to Bouton (1993, 1994), once a CS becomes ambiguous, the second learned information becomes context specific through a process called occasion setting. An *occasion setter* is a stimulus that helps differentiate an ambiguous stimulus by facilitating retrieval of a particular memory. For instance, in a typical occasion-setting procedure, a CS is followed by a US except when another stimulus (an occasion setter) is present. One of the special characteristics of occasion setting is that occasion setters do not typically impact responding directly. Instead, they clarify the meaning of an ambiguous stimulus by promoting retrieval of the appropriate experience. Bouton (1993, 1994) suggests that in order for the memory of extinction to be retrieved, the animal must be exposed to both the CS *and* the context in which extinction was carried out. In an extinction procedure, the extinction context acts as an occasion setter that facilitates recall of the memory of extinction. If the CS is presented outside of the extinction context, then the inhibitory association will not be retrieved, thereby resulting in a return of conditioned responding based upon the original memory of the CS being followed by the US. Based on this view, all second learned information about a stimulus is controlled in this manner.

Clinical Implications of the Renewal Effect

One of the reasons why the renewal effect has attracted substantial interest from researchers of animal learning as well as clinical psychologists is because of its considerable implications for the treatment of psychological disorders. Consider a person with a fear of snakes. Phobias are intense, irrational fears (irrational because the perceived threat is much greater than the actual threat) that cause marked impairment in the ability of an individual to function in his or her daily life. The behavioral treatment of choice is exposure therapy, which ultimately consists of exposure to the feared object or situation in the absence of any sort of traumatic outcome (i.e., extinction). In everyday life, psychological treatment is most commonly carried out in a therapist's office. During therapy, the client with a fear of snakes is gradually exposed to snakes until the client no longer feels anxious. At the conclusion of therapy, it is hoped that once the client returns to normal, everyday life, he or she will no longer experience substantial fear in the presence of snakes. Such a treatment plan is similar to the ABA renewal studies already described, which suggests that although the treatment might work in the therapist's office, there is a good chance that the client's fear will return once he or she encounters a snake outside of the therapist's office.

Mystkowski, Craske, and Echiverri (2002) have provided evidence that renewal of fear can occur following seemingly successful treatment of a fear of spiders. In their experiment, people with a fear of spiders showed a return of fear when they encountered a spider in a different place from where their treatment was provided. Findings such as these have resulted in research designed to study whether renewal of conditioned responding can be prevented. Given that extinction only holds up in the place where it was provided, one might consider providing treatment in the environment in which the client is most likely to encounter the fearful stimulus. This strategy is likely to prove more effective given that those environments should help retrieve the memory of therapy, thereby reducing the client's fear.

Another potential strategy for reducing the renewal effect is treatment in multiple contexts. Gunther, Denniston, and Miller (1998) provided fear conditioning to rats in one context, followed by extinction in a second context. When subjects were tested with the extinguished CS in a new context (an ABC renewal design), the typical return of fear was observed. However, if extinction was divided over three different contexts (B, C, and D), less renewal of fear was observed when testing was conducted in a new context. Findings such as these suggest that providing treatment in multiple settings might help reduce relapse in phobic individuals.

So far, I have discussed context only as the physical environment of an animal. However, many other cues can function as contexts. For instance, physiological states triggered by drugs or emotions can have effects similar to those described above. Anecdotally, when you are feeling down, you are more likely to recall a variety of negative experiences from your life, rather than your successes and other "happy thoughts." Emotional states can act as retrieval cues for other memories. Researchers have studied effects of this nature in the laboratory using drugs to induce different physiological states. Bouton, Kenney, and Rosengard (1990) provided rats with fear conditioning and then extinguished the animals when they were under the influence of an anti anxiety drug (Valium). When subjects were tested for fear once they were no longer under the influence of the drug, the researchers observed a return of fear. This finding is particularly troublesome given that clients are often prescribed medications to reduce anxiety while they undergo treatment of their fear. Once the client is anxiety free and medication is terminated, a return of fear is possible, given the change in physiological state. Studies using human participants provide support for this possibility. Mystkowski, Mineka, Vernon, and Zinbarg (2003) provided extinction treatment to college students while they were under the influence of either caffeine or a placebo in order to reduce their fear of spiders. When the researchers tested the students in a different physiological state than that in which treatment was provided, a greater return of fear was observed. Such effects have potentially profound implications for the treatment of conditioned fear and will hopefully lead to more promising outcomes for clients as the implications of these studies become better appreciated.

CUE COMPETITION

The Blocking Effect

For many years, researchers of Pavlovian learning assumed that conditioning was a relatively automatic process, provided that the CS and US occurred in a consistent manner close together in time (i.e., with a strong contingency and contiguity, respectively). However, in the late 1960s this view was seriously challenged by a group of phenomena known as cue competition. *Cue competition* refers to a family of effects in which the ability of a stimulus to produce a conditioned response following conditioning is impaired by the presence of other conditioned stimuli. That is, CSs (cues) appear to compete with one another for the role of predicting a US. One of the most thoroughly studied examples of cue competition is that of *blocking* (Kamin, 1969). Kamin provided rats with fear conditioning in which a noise was followed by shock. Once the noise was well conditioned, subjects then received additional training in which the noise and a new stimulus (a light) were again paired with shock multiple times. For ease of presentation (so that you don't have to keep track of which stimulus is which), the noise will be referred to as CS A and the light will be referred to as CS X (the target of testing). When subjects were subsequently tested with the CS X, no fear was observed. Kamin claimed that the prior conditioning of the tone "blocked" learning about the CS X. This result was particularly surprising, given the numerous pairings of CS X (and CS A) with shock, and seriously questioned whether contiguity alone is sufficient for learning.

Toward explaining the blocking effect, Kamin (1969) claimed that Pavlovian learning will only occur when an animal is surprised by the events in its environment. Specifically, Kamin stated that if an animal is surprised by a US, then this will stimulate the learning process, essentially causing the animal to search for causes of the surprising event. Applied to blocking, when the animal first receives CS A followed by shock, the presentation of shock would be unexpected (shocking, if you will), thereby activating the learning process and resulting in conditioning of CS A as a predictor of shock. Following sufficient training, the shock should no longer be surprising, as it is fully predicted by CS A. Now consider the next part of the blocking study, in which both the CS A and CS X are paired with shock. When CSs A and X are presented, the animal should expect the shock, thereby making it less surprising once it is received. This absence of surprise should greatly limit learning about the added stimulus (CS X). In essence, CS X provides no new information to the animal (it is redundant) and is blocked from the learning process.

The Rescorla-Wagner Model

Kamin's (1969) analysis of blocking truly revolutionized theories of Pavlovian learning for decades to come. Rescorla and Wagner (1972) developed a mathematical model of learning based in large part on Kamin's idea of surprise. In their model, surprise is viewed as a discrepancy between what an animal actually receives and what the animal expects to receive—the greater the difference, the greater the learning that will occur in a given conditioning trial. Essentially, learning reflects the difference between the size (or magnitude) of the US received and the magnitude of the US expected (based upon the CSs present on a given conditioning trial; i.e., US received minus US expected). Applied to blocking, the first stimulus, CS A, gains associative strength when the unexpected shock occurs because the US received is much greater than the US that is expected. As conditioning progresses, surprise (the difference between the US received and the US that is now predicted by CS A) is reduced as CS A becomes a better predictor of shock. When CS X is added to the conditioning process in the second part of the blocking procedure, surprise is reduced by CS A, thereby limiting learning about CS X. Hence, according to the Rescorla-Wagner model, the absence of responding to the CS X during testing reflects a learning deficit that cannot be reversed unless CS X is directly paired with the US in the absence of the previously conditioning CS A (as the shock would now be surprising when CS A is no longer present to predict it).

The US-Preexposure Effect

Cue competition can also occur between a CS and the context in which it is trained. One example of this is the US-preexposure effect (Randich & LoLordo, 1979), in which an animal receives exposure to the US prior to Pavlovian conditioning training. In a typical US-preexposure procedure, an animal is placed into a particular type of experimental chamber (context) and is presented with the US (e.g., shock) multiple times. Following preexposure to the US, conditioning trials are provided in which a CS (X) is paired with a US (e.g., CS X→shock). At test, subjects that received preexposure to the US are less fearful of CS X than are control subjects that did not experience preexposure to the US (i.e., that merely received X→shock pairings).

One potential explanation of the US-preexposure effect is that the animal develops a fear of the conditioning context as a result of its being associated with the shock. The Rescorla-Wagner model (1972) describes this as a consequence of the shock being unexpected (i.e., surprising), thereby allowing the context to become associated with the shock. Once the context becomes a good predictor of shock, surprise is reduced. Now consider the conditioning trials in which CS X is paired with shock. When the animal is placed into the context, the prior conditioning should generate an expectation of shock, such that when CS X is later paired with shock, surprise is reduced and learning about CS X is diminished. (Note the similarity between this explanation and the one provided for the blocking effect—US preexposure is essentially blocking by the context.)

A theoretically related effect to that of US preexposure is the effect of trial spacing on conditioning. Massed training in which CS-US pairings occur close together in time tends to produce less effective conditioning than does spacing the same number of conditioning trials over time (Gibbon, Baldock, Locurto, Gold, & Terrace, 1977). Essentially, spreading learning over time, as opposed to "cramming," leads to more effective learning. Again, the context is thought to play a role in this effect. From the perspective of the Rescorla-Wagner model (1972), both the CS and the context become good predictors of the US, given the reduced time between conditioning trials. In this instance, the context functions like a CS, thereby allowing learning about both the context and the CS, as they are both potential predictors of the surprising US. However, the context and the CS must share their prediction of the US, leaving the CS with less of an association than if the trials had been spread out in time. With spaced training, the context is present for some time between conditioning trials, which should result in an expectation of the US that is not presented. The absence of the US when it is otherwise expected, based upon the context, should result in extinction of the context so that the animal no longer expects the US except when the CS is present.

Comparator Theories

The Rescorla-Wagner model (1972) has been one of the most successful theories of Pavlovian learning. The reason for its success is not necessarily its accuracy, but rather its ability to inspire research and new theories of conditioning. One family of theories that followed the Rescorla-Wagner model is *comparator theories* (e.g., Gibbon & Balsam, 1981; Miller & Matzel, 1988), which view conditioning as a comparison of the predictive ability of the CS to that of other stimuli. When the CS is a relatively better predictor of the US than are other stimuli, conditioned responding will be observed. Scalar Expectancy Theory (Gibbon & Balsam, 1981) views this comparison as taking place between the CS and the context in which training is provided. According to this theory, two time intervals are compared—the time between the CS and the US (known as the trial time, T) and the time between conditioning trials (known as the cycle time, C). When the ratio of C/T exceeds some specified value, conditioned responding emerges. Applied to the effect of trial spacing, spacing conditioning trials adds more time to the trial time and thereby inflates the value of the numerator of this comparison (above the threshold for responding). Conversely, massed conditioning trials reduces the cycle time and consequently the ratio, thereby impairing conditioned responding. Notably, learning of these time intervals is not impaired by massed training. Instead, conditioned responding simply fails to emerge because the context is essentially too good of a predictor of the US.

The account of the trial spacing effect provided by Scalar Expectancy Theory (Gibbon & Balsam, 1981) differs from that provided by the Rescorla-Wagner (1972) model. Whereas the Rescorla-Wagner model views all response deficits as a consequence of learning impairments in which the CS does not readily condition due to the presence of other stimuli, Scalar Expectancy Theory attributes the lack of responding to a performance deficit resulting from the C/T comparison. Central to this theoretical difference is whether these response deficits are due to performance variables rather than to deficits in learning (as posited by the Rescorla-Wagner model). Efforts to test these competing views have provided US preexposure training prior to conditioning (Matzel, Brown, & Miller, 1987). Once the typical US-preexposure effect was observed, further training was provided in which the context was extinguished by exposing the animals to the context without any CS or US presentations. This training should increase the time between trials (the cycle time) and thereby increase the C/T ratio. Once this training was provided, the researchers observed greater conditioned responding to the CS, as would be anticipated by Scalar Expectancy Theory.

The Comparator Hypothesis

The comparator hypothesis (Miller & Matzel, 1988) is related to scalar expectancy theory (Gibbon & Balsam, 1981) in numerous ways, except, primarily, for the nature of the comparison. The comparator hypothesis claims that learning occurs whenever two or more stimuli occur close together in time (and space). Note that this view differs from the laws of learning discussed earlier that claimed that contiguity alone was not sufficient for learning (instead, factors such as surprise were also critical). However, according to the comparator hypothesis, performance is based upon relative associative strengths. What does this mean? Take, for example, a typical conditioning procedure in which a CS is paired with the US. According to the comparator hypothesis, three associations will be formed: (a) between the CS and the US, (b) between the CS and the context (the place in which the CS is trained), and (c) between the context and the US. At test, the CS is presented by itself, and the animal compares the associative strength of the CS to other stimuli that were present during *training* with the CS (the so-called comparator stimuli). The comparison works in the following manner: When the CS is presented at test, it activates memories (representations) of all of the stimuli that were present during its training, including the context. When the CS activates a memory of the context, the context can now (potentially) activate a memory of other stimuli that were present in that context (e.g., the US). Thus, the CS can directly activate a memory of the US, but it can also indirectly activate a memory of the US through the comparator stimulus (the context).

The comparator hypothesis claims that animals compare the strength of the US representation activated directly by the CS (association 1) to the strength of the US representation that is indirectly activated by the CS (through the comparator stimulus). This latter association will only be activated when the association between the CS and the comparator stimulus and between the comparator stimulus and the US are both strong. If either association is relatively weak, then the indirectly activated memory of the US will not be activated. Hence, in order to see a strong conditioned response to a CS, it must have a relatively strong association to the US (compared to other possible predictors of the US). Essentially, the CS needs to be a better predictor of the US than any other stimulus. This can potentially be accomplished in a couple of ways. The first way is if the CS has no comparator stimulus, but this isn't practical in that training must take place somewhere (i.e., in some type of context). The second way is if the training stimulus's comparator stimulus is only weakly associated with the US. Applied to a typical acquisition procedure—a CS is paired with the US. This would allow the CS to become associated with the US as well as with the context (because they are present at the same time—i.e., with good contiguity). At the same time, the context would become associated with the US (again, due to good contiguity). However, if there is a relatively long period of time between conditioning trials, the animals will be exposed to the context without the US. This would result in a weakening (extinction) of the context-US association and therefore a relatively ineffective comparator stimulus.

Now consider an instance in which conditioning is conducted with close spacing between trials. Yin, Barnet, and Miller (1994) provided rats with fear conditioning training in which a CS was trained in either a massed or spaced manner (i.e., trials were presented either close together or spread out in time). The comparator hypothesis would predict that the CS-US association would be the same in both conditions, as the CS and US are presented together. Similarly, the association between the CS and the context would also be equivalent across both groups because the CS and the context are paired the same number of times. What should differ between groups is the strength of the context-US association, in which this association would be expected to be stronger in the massed group than in the spaced group because the context-US association would extinguish more in the spaced group (due to the greater time between trials). As a result, the memory of the US activated through the comparator stimulus (the context) should be greater in Group Massed than in Group Spaced, thereby leading to less responding in Group Massed. This is what Yin et al. found—strong conditioned responding in Group Spaced and weak conditioned responding in Group Massed (i.e., the trial spacing effect). What they did next was even more interesting. They then extinguished the context for both groups by placing the animals in the context without giving them the US. Note that this should weaken the context-US association for the massed group, thereby leading to a less effective comparator stimulus. At test, they found that context extinction resulted in strong conditioned responding in both groups. That is, the trial spacing effect went away. Interestingly, this suggests that the trial spacing effect doesn't influence "learning"; instead, it influences performance.

The comparator hypothesis can also explain the blocking effect that was originally observed by Kamin (1969). Recall that in a typical blocking procedure, a CS (A) is paired with the US in the first stage of training and then CS A (the blocking stimulus) and a new stimulus, CS X (the blocked stimulus), are paired with the US in the second stage of training. At test, conditioned responding to the blocked stimulus (X) is weak relative to a control condition lacking the critical first stage of training. The comparator hypothesis explains this effect as being due to a performance deficit resulting from the comparator process, not to a failure to learn an association between CS X and the US. That is, animals learn an X-US association, but they also learn associations between X and A and between A and the US. These strong X-A and A-US associations result in decreased performance to CS X (again, CS A is too good of a predictor of the US). The comparator hypothesis makes the prediction that weakening the A-US association through extinction of CS A should result in a weakened memory of the US activated through the comparator stimulus (A), and consequently an increase in conditioned responding to CS X. Blaisdell, Gunther, and Miller (1999) confirmed this prediction when they provided rats with blocking training followed by extinction of CS A and observed an increase in responding to CS X. These findings suggest that animals learned about the blocked stimulus (X) but failed to respond to it due to its strong associations to the blocking stimulus (A), which in turn is associated with the US.

SUMMARY

Over the past several decades, researchers have learned much about the nature and function of Pavlovian learning. These developments have their roots in the classic work of Pavlov (1927) and have extended his work to the current state of knowledge, which points to the rich nature of Pavlovian learning. Pavlovian conditioning is a robust form of learning that provides knowledge to animals central to their survival. Through the mechanisms of Pavlovian learning, a wide variety of knowledge about our world can be acquired and used to understand not only what will occur in our environment, but also when these events will occur. Future research will continue to analyze the critical determinants of learning and performance of Pavlovian associations.

REFERENCES AND FURTHER READINGS

Barnet, R. C., & Miller, R. R. (1996). Temporal encoding as a determinant of inhibitory control. *Learning and Motivation, 27*, 73–91.

Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking through deflation of the block stimulus. *Animal Learning & Behavior, 27*, 63–76.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114*, 80–99.

Bouton, M. E. (1994). Context, ambiguity, and classical conditioning. *Current Directions in Psychological Science, 3*, 49–53.

Bouton, M. E. (1997). Signals for whether versus when an event will occur. In M. E. Bouton & M. S. Fanselow (Eds.), *Learning, motivation, and cognition: The functional behaviorism of Robert C. Bolles* (pp. 385–409). Washington, DC: American Psychological Association.

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10*, 445–466.

Bouton, M. E., Kenney, F. A., & Rosengard, C. (1990). State-dependent fear extinction with two benzodiazepine tranquilizers. *Behavioral Neuroscience, 104*, 44–55.

Brogden, W. J. (1939). Sensory preconditioning. *Journal of Experimental Psychology, 25*, 323–332.

Gibbon, J., Baldock, M. D., Locurto, C. M., Gold, L., & Terrace, H. S. (1977). Trial and intertrial duration in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes, 3*, 264–284.

Gibbon, J., & Balsam, P. (1981). Spreading association in time. In C. M. Locurto, H. S. Terrace, & J. Gibbon (Eds.), *Autoshaping and conditioning theory* (pp. 219–253). New York: Academic Press.

Gunther, L. M., Denniston, J. C., & Miller, R. R. (1998). Conducting exposure treatment in multiple contexts can prevent relapse. *Behaviour Research and Therapy, 36*, 75–91.

Hume, D. (1964). *Treatise of human nature* (L. A. Selby-Bigge, Ed.). London: Oxford University Press. (Original work published 1739)

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & M. R. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Konorski, J., & Szwejkowska, G. (1950). Chronic extinction and restoration of conditioned reflexes. I. Extinction against the excitatory background. *Acta Biologiae Experimentalis, 15*, 155–170.

Konorski, J., & Szwejkowska, G. (1952). Chronic extinction and restoration of conditioned reflexes. III. Defensive motor reflexes. *Acta Biologiae Experimentalis, 16*, 91–94.

Matzel, L. D., Brown, A. M., & Miller, R. R. (1987). Associative effects of US preexposure: Modulation of conditioned responding by an excitatory training context. *Journal of Experimental Psychology: Animal Behavior Processes, 13*, 65–72.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51–92). San Diego, CA: Academic Press.

Mystkowski, J. L., Craske, M. G., & Echiverri, A. M. (2002). Treatment context and return of fear in spider phobia. *Behavior Therapy, 33*, 399–416.

Mystkowski, J. L., Mineka, S., Vernon, L. L., & Zinbarg, R. E. (2003). Changes in caffeine states enhance return of fear in spider phobia. *Journal of Consulting and Clinical Psychology, 71*, 243–250.

Nelson, J. B. (2002). Context specificity of excitation and inhibition in ambiguous stimuli. *Learning and Motivation, 33*, 284–310.

Pavlov, I. P. (1927). Conditioned reflexes. London: Oxford University Press.

Randich, A., & LoLordo, V. M. (1979). Preconditioning exposure to the unconditioned stimulus affects the acquisition of a conditioned emotional response. *Learning and Motivation, 10*, 245–277.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin, 72*, 77–94.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning. II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rizley, R. C., & Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative and Physiological Psychology, 81*, 1–11.

Rosas, J. M., & Alonso, G. (1997). Forgetting of the CS duration in rats: The role of retention interval and training level. *Learning and Motivation*, *28,* 404–423.

Yin, H., Barnet, R. C., & Miller, R. R. (1994). Trial spacing and trial distribution effects in Pavlovian conditioning: Contributions of a comparator mechanism. *Journal of Experimental Psychology: Animal Behavior Processes, 20*, 123–134.

35

Taste-Aversion Learning

W. Robert Batsell
Kalamazoo College

In *taste-aversion learning*, an organism consumes a substance and experiences a nausea-inducing event. When this organism subsequently reencounters the edible substance, it will consume significantly less of it, especially when compared to controls that experienced the edible, but not the illness. Hence, the organism learns a taste aversion. Thus, taste-aversion learning is highly adaptive because it enables the foraging organism to identify potential sources of illness and learn to avoid them in the future. Considering its adaptive value, it is not surprising that evidence of taste-aversion learning has been shown across the entire animal kingdom including mammals, birds, reptiles, amphibians, fish, and insects. It is worth noting that the terms *flavor-aversion learning* and *taste-aversion learning* are used interchangeably, but technically, these terms are different. Taste-aversion learning should refer specifically to situations in which the edible has only a taste component, whereas flavor-aversion learning refers to situations in which the edible has both taste *and* odor components. Thus, most taste-aversion learning situations are more accurately flavor-aversion learning situations.

Taste-aversion learning is a form of classical conditioning (see Classical Conditioning, Chapter 33 in this volume) in which the edible is the conditioned stimulus (CS) and the illness-inducing stimulus or event is the unconditioned stimulus (US). Illness would be the unconditioned response (UCR), and either illness or an unwillingness to consume the edible could be the conditioned response (CR). As we will see in the Theory section, taste-aversion learning has often been viewed as somewhat different from other forms of classical conditioning. One example of this proposed difference is that in most forms of classical conditioning, the CS is a relatively neutral cue that elicits a weak response at best. Yet, in taste-aversion learning, the CS is often a food substance that could serve as a US. Moreover, this food substance can have reinforcing properties for the hungry or thirsty organism, so that taste-aversion learning may appear to be a *counter-conditioning procedure* in which an appetitive substance is now a signal for aversive consequences.

METHODS

Conditioning

The most common taste-aversion experiments utilize rats as subjects because they readily learn taste aversions, are easily domesticated, can be housed in large numbers, are relatively inexpensive, and do not vomit. In these experiments, the rat drinks a flavored solution, such as sweet saccharin. For optimal conditioning, it is important to limit the amount of solution (5 to 10 ml) and the time of exposure to the solution (5 to 10 min) because increasing exposure to the CS will decrease the strength of the taste aversion. The choice of the CS flavor is less important for optimal conditioning because profound aversions occur with tastes that are sweet (e.g., saccharin), sour (e.g., hydrochloric acid), bitter (e.g., denatonium), and salty (e.g., saline).

Similarly, taste aversions occur with a wide range of USs. Although the most common experimental means to induce illness is an intraperatoneal injection (to the stomach) of the mild emetic lithium chloride, taste aversions occur with a range of illness-inducing agents including irradiation (e.g., Garcia, Kimeldorf, & Koelling, 1955; Smith & Roll, 1967) and rotational stimulation (e.g., Batsell & Pritchett, 1995; Braun & McIntosh, 1973). Regardless of the type of illness-inducing US, aversion strength increases along with the intensity of the US.

As in other classical conditioning experiments, an important experimental variable is the CS-US interval (i.e., the amount of time that elapses between termination of the CS and presentation of the US). Interestingly, the strongest taste aversions appear to be produced when the CS-US interval is a shorter time period, such as 15 minutes; however, reliable taste aversions can occur with backward conditioning (Barker, Smith, & Suarez, 1977) and extended CS-US intervals (e.g., Garcia, Ervin, & Koelling, 1966). Examination of taste-aversion learning with a long CS-US interval has been a key theoretical issue in the study of taste-aversion learning, and it will be addressed in the Theory section.

Testing

There are several testing procedures that insure accurate detection of a taste aversion. First, testing can occur at any period following the initial taste and illness pairing, but early tests may be compromised. For example, nauseous organisms will reduce their consumption of even familiar and preferred edibles, so tests conducted while the rat is still recovering from illness would confound detection of a taste aversion. Furthermore, research has shown that taste aversions may be artificially weak up to 24 hours following conditioning due to nonassociative issues (e.g., Batsell & Best, 1992), but aversion strength stabilizes by 48 hours after conditioning (Batsell & Best, 1994). Therefore, a retention interval of three or more days will allow for accurate recording of the taste aversion.

Second, the experimenter should consider the benefits of using either a one-bottle or a two-bottle test. In the typical one-bottle test, the deprived rat is given limited access to the target solution (e.g., a 10-minute or 20-minute exposure). This test essentially pits the rat's thirst and motivation to consume liquid against its unwillingness to consume the target solution. The rat's consumption of the target fluid is a reflection of aversion strength. Past experiments have confirmed that this test is an effective means of recording taste aversions, and it is particularly suited when comparing groups that have aversions of differential strength (Batsell & Best, 1993). Conversely, the two-bottle test is a choice test in which the rat is given simultaneous access to the target solution in one bottle and a different flavor in a second bottle. In this test, the aversion is the percentage of the target fluid consumed relative to the total amount consumed. The two-bottle test is particularly suited for the detection of very weak aversions (e.g., Dragoin, R. T. McCleary, & G. E. McCleary, 1971; Grote & Brown, 1971). One consideration that is crucial to the interpretation of the two-bottle test is the choice of the comparison flavor. If the comparison flavor is water or a familiar, preferred flavor, the test may be biased by preference for the comparison flavor rather than an aversion to the target fluid. As a result, between-group differences in aversion strength are obscured. Similarly, if the comparison flavor is novel, the rat's natural unwillingness to consume a novel substance (*neophobia)* may compromise the validity of the test. Therefore, the experimenter's choice of the comparison flavor is a critical decision that can influence the accuracy of the two-bottle test.

Third, although the consumption test is the most common test, Grill and Norgren's (1978) taste reactivity test can also provide valuable data. In this test, the experimenter records the organism's response to the edible, often via videotape, and notes characteristic disgust responses. Some of these characteristic disgust responses would include gaping (the organism extends its tongue out of its mouth), chin rubbing (the organism vigorously paws at its mouth), and head shaking (the organism moves its head side to side). It is worth noting that some researchers make the distinction between taste aversion and taste avoidance (e.g., Parker, 2003). The behavior most commonly used as evidence of flavor-aversion learning is consumption of the edible, but Parker and others use the term *taste avoidance* for these scenarios in which the organism avoids consumption of the edible. Instead, they use the term *taste aversion* to refer to the set of negative behaviors from the taste reactivity test that provide evidence that the organism now dislikes the edible.

THEORY

John Garcia and the Advent of Taste-Aversion Learning Research

The demonstration of taste-aversion learning across the animal kingdom suggests that the phenomenon is millions of years old, but the development of a laboratory preparation to investigate taste-aversion learning did not originate until the 1950s with the work of John Garcia and his colleagues. Garcia's academic journey is as fascinating as his research accomplishments. Garcia was the child of Hispanic migrant workers, who earned their living following the fruit seasons throughout California. Thus, the young Garcia grew up in the fields of California, observing the behavior of the family's pets and natural wildlife. This "real-world" education was in sharp contrast to his more formal elementary education. Because Garcia primarily spoke Spanish, he had difficulty adapting to the English classroom, and his third-grade teacher remarked that he would never amount to much (Garcia, n.d.). Later, Garcia served in World War II, and was able to attend college on the recently created G.I. Bill. He chose to attend the University of California, Berkeley, where he advanced to complete his graduate studies in the lab of E. C. Tolman. To appreciate Garcia's unique perspective, it is helpful to recall that the predominant theoretical psychological perspective during the late 1940s was neobehaviorism. As such, an animal like the rat was viewed as a machine of habits and twitches, shaped by any events that it encountered. Onto this theoretical landscape, Garcia brought his background of observing animals in their natural habitat as they faced ecological pressures.

After Garcia earned his graduate degree in 1950, he worked for the U.S. Navy, and it was here that he began taste-aversion learning research. In the early preparations, Garcia, Kimelolorf, and Koelling (1955) used radiation to induce illness. In this preparation, rats drank a highly preferred saccharin solution or water while confined in the experimental chamber. Rats were then exposed to one of two radiation conditions for a 6-hour period; the 30-r group (r = roentgen, which is a measure of radiation) received 5.0 r/hr whereas the 57-r group received a stronger 9.5 r/hr exposure. Control groups also drank water or saccharin, but a lead shield protected them from radiation. Two-bottle testing commenced two days later as rats were given a choice of water or saccharin in their home cage. The control rats continued to show the high preference for saccharin that they displayed before conditioning (86 percent preference of saccharin over water). On the initial tests, the 30-r group showed a 35 percent preference for saccharin and the 57-r group's saccharin preference was less than 5 percent. These data were the first to demonstrate empirically that an illness-inducing agent could produce a dramatic decrease in consumption of a preferred flavor. Interestingly, Garcia et al. extended testing up to 60 days after the single conditioning trial, and despite these multiple exposures to saccharin during testing, Group 57-r never returned to their preconditioning saccharin preference level. Indeed, the saccharin preference level of Group 57-r hovered around 70 percent for the 10 days of testing. These data confirm the long-lasting nature of taste-aversion learning and its relative resistance to extinction.

Over the next 30 years, Garcia would continue to be at the forefront of taste-aversion research while working at the University of California, Los Angeles. During this time, he would examine phenomena such as cue-to-consequence learning, long-delay learning, and potentiation, all of which created intellectual commotion because they were not anticipated by previous research. Nonetheless, in each case, Garcia's findings were replicated, and even if his theoretical treatments did not always stand the test of time and future work, his initial contributions stimulated considerable research and thought.

Is Taste-Aversion Learning a Unique Form of Learning?

In the late 1960s and early 1970s, Garcia's research motivated many researchers to conduct taste-aversion research to replicate and extend his novel findings. Indeed, from a historical standpoint, one could reasonably conclude that the 1970s and the early 1980s were the heyday of taste-aversion research, as the leading learning journals included many taste-aversion learning articles and many definitive collections appeared (Barker, Best, & Domjan, 1977; Braveman & Bronstein, 1985).

Cue-to-Consequence Learning

To appreciate the historical import of the introduction of cue-to-consequence learning, one must consider the prevailing mindset in the 1960s. At this time, a concept known as the general process theory of learning provided a theoretical framework to explain most learning phenomena. According to this theory, the nervous system of living organisms was designed to learn that basic stimuli (CS) could signal the presence or absence of biologically significant events (the US)—a statement that would be accepted by almost all researchers today. Yet in the 1960s, this proposition would have included within this readiness to learn that *any* neutral stimulus such as a light, tone, or taste could be equally effective as a CS, and *any* biologically significant stimulus such as a shock or illness could be equally effective as a US. Hence, assuming that the relative intensities of the CSs and USs were equal, any combination of these stimuli should produce equivalent levels of learning.

Garcia and Koelling (1966) tested this hypothesis in an experiment that is often referred to as the "bright-noisy water experiment." In this experiment, they used two types of CSs (taste and audiovisual cues) and two types of USs (shock and sickness). The rats were allowed to drink at a spout that contained a flavored liquid (saccharin); when the rat's tongue touched the spout it activated the flash of a light along with a clicking sound (i.e., bright-and-noisy water). After the animals had consumed the liquid, half of each group received an injection of poison while the other half received shock. For testing, the taste component was separated from the audiovisual component so the rats were tested either for their aversion to drinking the taste of saccharin or to drinking water in the presence of the light and noise. The rats that experienced the illness US drank significantly less of the taste CS than of water in the presence of the light and tone. In contrast, the rats that experienced the shock US drank significantly less of the water in the presence of the light and noise than of the saccharin taste. This novel finding was in direct conflict with the general process rheory of learning: Certain cues were better associated with certain consequences (e.g., taste with illness, audiovisual cues with shock), hence cue-to-consequence learning. In light of the influence of this work on learning theory, an interesting historical footnote is that Garcia and Koelling had a difficult time publishing this research; in fact, they submitted the manuscript to multiple academic journals before it was accepted for publication (Garcia, n.d.).

Thus, Garcia and Koelling (1966) demonstrated the principle of cue-to-consequence learning in that certain combinations of stimuli are more easily associated than other combinations (taste with illness better than taste with shock). Garcia and Koelling argued that cue-to-consequence learning reflects a genetic predisposition for the selective association of certain combinations of CSs and USs. It is important to note that pairing a taste with shock can produce an effective taste aversion, but it is not as effective as pairing taste with illness. Later work from other labs replicated the basic cue-to-consequence finding (Domjan & Wilson, 1972). For a period of time it was thought that this effect was specific to taste-aversion conditioning; however, the presence of selective associations occur in fear learning (e.g., Ohman & Mineka, 2001) and in instrumental conditioning (e.g., LoLordo & Droungas, 1989).

One-Trial Learning and Long-Delay Learning

The most common preparations for studying classical conditioning (i.e., fear conditioning, autoshaping, eyeblink conditioning) all share similar procedural constraints in that they require multiple CS-US pairings to elicit a CR, and extending the CS-US interval—even if it is only a matter of seconds—is sufficient to weaken learning. In contrast, reliable taste aversions are produced after only a single conditioning trial and after extended CS-US intervals (e.g., Garcia, Ervin, & Koelling, 1966). Garcia's work in long-delay learning spurred much subsequent research. For example, Smith and Roll (1967) conducted saccharin-radiation conditioning, but they varied the delay interval between saccharin consumption and radiation exposure (i.e., the CS-US interval), including delays of 0.0 hour, 0.5 hour, 1 hour, 2 hour, 3 hour, 6 hour, 12 hour, and 24 hour. Saccharin preference testing occurred one day later. Reliable taste aversions occurred even with a 6-hour or 12-hour delay imposed between taste consumption and illness induction. Although the initial interpretation was that one-trial learning and long-delay learning may be unique to taste-aversion learning, subsequent research has confirmed that one-trial learning can be observed in other preparations, and that animals can learn across extended CS-US intervals if interfering stimuli are kept at a minimum (for a review, see Bouton, 2006).

Synergistic Conditioning

In 1927 Pavlov described the effect of pairing a weaker CS and a stronger CS with a single US. During testing, Pavlov observed that learning was significantly stronger to the stronger CS than the weaker CS. In fact, learning to the weaker CS was significantly less than if only the weak CS was paired with the US. Pavlov coined the term "overshadowing" to describe this outcome as the presence of the strong CS competed with the weak CS to decrease learning to the latter stimulus. Evidence of overshadowing has since been reported in different classical conditioning preparations (e.g., Mackintosh, 1976).

In the late 1960s and early 1970s, research in other areas of classical conditioning was providing evidence of the rules of associative learning—in particular, the rules that may operate during compound conditioning. Leon Kamin introduced another example of competitive conditioning in 1969 when he reported the results of blocking in fear-conditioning experiments. In these experiments one CS (a tone) was presented with an aversive US (shock) for multiple trials. In the second conditioning phase, the pretrained CS (tone) and a novel CS (a light) now signaled the shock US. During testing, Kamin observed a strong fear CR to the pretrained tone, but no fear response to the light. He concluded that the pretraining of tone-shock prevented or "blocked" learning to the light because the light presented no new information. Overshadowing and blocking both suggest that cues compete with one another for associative strength; this finding led theorists to incorporate rules of competitive conditioning into formal models of associative learning (cf. Rescorla & Wagner, 1972).

Once again, Garcia's work would present a finding that would challenge the assumption that cues must compete with one another. In 1979 Rusinak, Hankins, Garcia, and Brett published a series of experiments that demonstrated a new learning phenomenon, *taste-potentiated odor aversion* (TPOA). In these studies, the experimental group drank a strong taste plus a weak odor compound prior to illness induction, whereas the control group was given only the weak odor-illness pairing. During odor testing, the experimental group had a significantly stronger odor aversion than the control group. In other words, instead of the strong taste overshadowing the weak odor, the presence of the taste strengthened or "potentiated" the odor aversion. The phenomenon of TPOA immediately attracted the attention of researchers because it was in direct opposition to the principles of competitive conditioning, and thus it refuted predictions derived from formal models of learning.

Subsequent research replicated TPOA and established that it was the opposite of overshadowing and an example of synergistic conditioning, in which the presence of multiple cues do not compete for, but enhance, learning to other cues. In consideration that TPOA was the opposite of the competitive conditioning effect of overshadowing, it was speculated that a synergistic conditioning opposite to blocking could also exist, but it was 20 years before the reliability of such an effect was reported (Batsell & Batson, 1999; Batsell, Paschall, Gleason, & Batson, 2001; Batson & Batsell, 2000). In the first set of these studies (Batson & Batsell), experimental rats were pretrained with odor-illness in Phase 1 before they were given a pairing of taste plus odor-illness in Phase 2. This design is procedurally the same as Kamin's blocking design. Compared to rats that had only the taste plus odor-illness pairing, the experimental rats displayed a significantly stronger taste aversion, a surprising outcome because models that incorporate competitive conditioning would have predicted blocking of the taste aversion. In a later set of studies, Batsell et al. reported the phenomenon was symmetrical because taste pretraining before taste plus odor-illness conditioning produced a significantly stronger odor aversion. The authors have used the term "augmentation" to describe the symmetrical synergistic conditioning effect obtained in the blocking design. Although there are different theoretical interpretations of synergistic conditioning (Durlach & Rescorla, 1980; Garcia, Lasiter, Bermudez-Rattoni, & Deems, 1985), the exact mechanism underlying TPOA and augmentation remains to be determined (Batsell & Paschall, in press).

APPLICATIONS

Taste-aversion research is a fertile area that has generated countless experiments to understand basic associative learning principles, and these findings have also been applied to many human conditions. In the following sections, I will review how results from taste-aversion research have been applied to human food rejection, chemotherapy treatment, and alcohol-aversion therapy.

Taste-Aversion Learning and Human Food Rejection

Lab-based investigations of taste-aversion learning have also led to studies of food rejection in humans—many of which involve variations of taste-aversion learning—but some broader distinctions are necessary. An organism's rejection of a food can be unlearned or learned. *Dislike*, for example, is the organism's unlearned rejection of an edible based on its taste, odor, or texture (Rozin & Fallon, 1980). There are multiple types of learned food rejection, and the most commonly studied form of learned food rejection in both humans and nonhumans is taste-aversion learning. With humans, a number of retrospective surveys (e.g., Garb & Stunkard, 1974; Logue, 1985; Logue, Ophir, & Strauss, 1981) confirm that taste aversions are learned through classical conditioning. The use of retrospective questionnaires has involved a large population completing a survey related to food rejection, illness experience, food type, and so on. The results of these questionnaires confirm that taste-aversion learning is quite common, with the percentage of individuals reporting at least one taste aversion varying from a low of 38 percent (Garb & Stunkard) to a high of 84 percent (de Silva & Rachman, 1987). Moreover, the results have consistently indicated that these food aversions are produced via classical conditioning, and they show the aforementioned predominant features of taste-aversion learning such as long-delay learning, one-trial learning, and persistence across time (e.g., Garb & Stunkard; Logue, 1985; Logue et al., 1981; de Silva & Rachman).

Although taste-aversion learning due to classical conditioning is the most commonly studied form of food rejection, there are other types of learned food rejections in humans. Rozin and Fallon (1980) identified three other types of food rejection in humans: inappropriate, danger, and disgust. First, *inappropriate* food rejection occurs when a food is rejected because it has no nutritional value and is inorganic (e.g., sand). Second, food rejection may also be due to *danger*, when the individual learns the negative consequences of ingestion (i.e., poisonous mushrooms). Third, a taste rejection based on *disgust* occurs when the individual learns the nature or origin of the edible, and then refuses to consume it (e.g., learning that a consumed food was not chicken, but snake). Disgust and danger rejections share a common feature in that they are produced by ideational or cognitive factors, often in the absence of nausea.

At present, there have been only a few investigations of the cognitive factors that mediate food rejections (e.g., Batsell & A. S. Brown, 1998; Rozin, 1986). For example, Rozin reported that the pairing of a taste with a disgust-eliciting stimulus produces food rejection. In his analysis of one-trial shifts in taste preference, Rozin reported that 26.5 percent of taste dislikes were mediated by disgust. More recently, Batsell and A. S. Brown reported the results of a retrospective questionnaire that also identified cognitive factors in taste-aversion learning. Categories of these cognitive aversions included disgust, negative information, or forced consumption. Subsequent research confirmed the forced-consumption scenario was a unique situation that introduced a social/power dynamic with a to-be-consumed edible that may have been disliked or disgusting (Batsell, A. S. Brown, Ansfield, & Paschall, 2002). The crucial components of the forced-consumption scenario, as identified by the participants, were that (a) they were forced to do something against their will and (b) their protests went unheeded. Thus, the investigation of forced consumption shows the intertwining of social interaction, cognitive processes, food, and aversive consequences.

Chemotherapy and Treatment

As stated earlier, taste-aversion learning provides the foraging animal with an effective adaptation to prevent repeated samplings of poisonous foods. The effectiveness of this mechanism is evident in that a single experience with taste and illness is often sufficient for humans or animals to stop eating the tainted edible. Yet, in some modern situations, the advantage of taste-aversion learning actually becomes a disadvantage. The most dramatic contemporary example is the use of chemotherapy to treat cancer. Most common chemotherapy treatments often produce the side effect of pronounced illness, which may last hours or days after treatment. Research has reported that cancer patients identify nausea and vomiting as the most distressing side effect of treatment (e.g., Boakes, Tarrier, Barnes, & Tattersall, 1993). Even though the illness side effects have been greatly diminished by the advent of antinausea medication, many of the medications are inconsistent, and taste-aversion learning requires only one opportunity to occur.

Posttreatment Vomiting and Nausea

Technically, researchers in this area have identified two different illness periods that can influence the cancer patient's eating habits. The first is *posttreatment vomiting and nausea* (PVN), the nausea from the treatment itself (i.e., PVN is the unconditioned or unlearned response to the US of treatment). Obviously, any foods eaten before chemotherapy treatment or during this nausea period are possible targets for taste-aversion learning. In 1978 I. L. Bernstein conducted research to determine if gastrointestinal (GI) toxic chemotherapy produced taste aversions in humans. In this study, she tested whether children undergoing chemotherapy treatment would develop an aversion to a unique ice cream flavor. Children in the experimental group (CS-US) ate Mapletoff ice cream (a combination of black walnut and maple flavor extracts) before GI toxic chemotherapy, whereas the children in the control group (US alone) did not eat the ice cream before chemotherapy treatment. Testing occurred either two or four weeks later. During testing, group CS-US showed a significantly lower preference for the Mapletoff ice cream (21 percent) compared to group US alone (67 percent). The fact that group US alone did not show a weakened preference for the Mapletoff ice cream confirmed that the decreased consumption of group CS-US was a learned aversion specific to ice cream and not a generalized decrease in consumption because of malaise, which both groups experienced. One important side note to this research is that I. L. Bernstein noted that many of the participants were well aware that the illness they experienced arose from their treatment sessions; nonetheless, they still learned powerful associations between ice cream and illness.

In consideration of the central role of taste-aversion learning to these deleterious side effects, much research has been initiated to lessen this suffering of cancer patients. One successful application of the logic of taste-aversion learning research to chemotherapy research is the introduction of a scapegoat flavor (I. L. Bernstein, Webster, & I. D. Bernstein, 1982). The logic of the scapegoat technique arose from previous research that showed if Taste A and Taste B are presented in series before illness (Taste A→ Taste B→illness), the aversion to the more contiguous Taste B will be significantly stronger than the less contiguous Taste A (e.g., Revusky, Parker, & Coombes, 1977). I. L. Bernstein et al. were aware that many patients would consume their normal diet before experiencing their chemotherapy treatment, so even though a long delay may be present between eating and illness, long-delay learning is a hallmark of taste-aversion learning. Instead, they examined the effects of having the patient eat a novel food that was not part of his or her regular diet immediately before chemotherapy treatment. In this way, the patient's regular meal would be the distant Taste A and the scapegoat flavor would be the recent, novel Taste B. Three groups of children participated in this experiment. The control group was untreated, the experimental group had their normal diet before chemotherapy, and the scapegoat group was given the Mapletoff ice cream immediately before chemotherapy treatment. The dependent variable of interest was the number of aversions that developed to foods from the child's diet. Not surprisingly, the control group developed few food aversions as a result of their placebo treatment. The experimental group reported the highest number of food aversions that developed because of their chemotherapy treatment. Importantly, the scapegoat group reported significantly fewer taste aversions to their normal diet than the experimental group, and the number reported by the scapegoat group was not significantly higher than the control group. Thus, the use of the interfering scapegoat technique is an effective means to protect the cancer patient's normal diet.

Anticipatory Vomiting and Nausea

Anticipatory vomiting and nausea (AVN), the learned-aversion response resulting from chemotherapy treatment, occurs after the patient has experienced at least one bout of illness-producing treatment. Then, when they reencounter some of the cues that were experienced with illness, these cues will elicit nausea and vomiting CRs. The incidence rate of AVN has varied in a number of different studies; however, Morrow and Dobkin (1988) surveyed the literature and reported a median of 33 percent of patients experience AVN. Patients have reported a variety of cues can trigger AVN; the most common triggers of AVN, in order, were odors, "no particular event" and places/contexts (Boakes et al., 1993).

Linda Parker, on the basis of past taste-aversion research, has developed an animal model of AVN that may prove invaluable in generating new behavioral treatments for this disorder (e.g., Parker, Kwiatkowska, & Mechoulam, 2005). In Parker's model, she uses the musk shrew because it reliably shows vomiting and retching (pronounced opening of the mouth) in response to lithium-induced illness (recall that rats do not vomit). These shrews learn that a neutral cue, such as a context, can serve as a signal for illness. After three pairings of context-illness, returning the shrew to this context reliably increases retching behavior. Next, this model helps identify drugs that could reduce context-induced illness. Interestingly, Parker et al. found that the antiemetic drug ondansetron, which is effective in reducing PVN, had no effect on reducing conditioned retching responses. Encouragingly, cannabidiol (a nonintoxicating cannibinoid that occurs in marijuana) was very successful in eliminating the conditioned retching responses that are analogous to AVN. Both the mechanism of this effect and its ultimate applicability to reducing AVN in the human population remain to be determined, but the potential is promising.

Alcohol Aversion Therapy

In the Chemotherapy and Treatment section, it is clear how a medical treatment inadvertently activates taste-aversion learning, which produces a side effect that interferes with cancer treatment. The next application, alcohol aversion therapy, involves a situation in which the treatment involves direct activation of taste-aversion learning to *facilitate* treatment. The use of emetics to decrease alcohol consumption is actually quite old; in 1789 the pioneer American psychiatrist Benjamin Rush reported poisoning a drunkard's rum to force him off the drink.

Baker and Cannon (1979) provided one of the better case studies detailing the effectiveness of illness-inducing or emetic treatment for alcoholism. In this case, a male alcoholic drank the oral emetic syrup of ipecac along with his favorite liquor. Over the next 20 minutes, the alcoholic alternated between drinking alcohol, water, and regurgitating. After this 20-minute period, the patient was given a small amount of beer that contained potassium antimony tartrate to prolong the nausea. At this time, the patient was instructed to think of all the problems drinking had caused him. This conditioning procedure was repeated five times. Immediately after treatment, the patient's intake of alcohol decreased by 50 percent, and the patient's negative feelings toward alcohol increased. The experimenters monitored the patient's postdischarge drinking status and outpatient adjustment for nine months. For the first three months, the patient drank no alcohol at all. During the next four months, he drank on 29 days, he drank only beer, and he never consumed more than three beers a day. The patient reported that even though he drank beer, both wine and hard liquor remained unappealing. Eventually, the patient did fall off the wagon.

Although the evidence from this case study is suggestive of the effectiveness of alcohol-aversion therapy, a later study was even more convincing. Cannon, Baker, and Wehl (1981) randomly assigned male alcoholic volunteers to three treatment conditions: (a) emetic aversion therapy, (b) shock aversion therapy, and (c) control, which received no aversion therapy. All patients also participated in a multifaceted alcoholism inpatient program (e.g., group, individual, marital, and family therapy; assertion training; relaxation training; alcohol education; Alcoholics Anonymous; etc.). The emetic aversion treatment consisted of 5 sessions in which a variety of alcoholic flavors were followed by nausea and vomiting. The shock aversion treatment consisted of 10 sessions, in each of which the patient experienced a variety of alcoholic flavors paired with variable-intensity electric shocks. Control patients received neither aversion treatment. During testing, each alcoholic participated in a taste test that evaluated behaviors consistent with conditioned taste aversions: (a) moving the glass away from the face in disgust after sniffing but before sipping; (b) grimacing; (c) gagging, choking, or coughing; and (d) pushing the drink away after sipping. The observers also made a subjective appraisal of whether the patient appeared to dislike each drink.

The results showed that emetic aversion, but not shock aversion, produced pronounced aversion to alcohol across all measures. Compared with the control group, alcoholics who received emetic aversion conditioning exhibited greater increases in heart rate to alcohol flavors, drank less during the taste test, reported more negative attitudes toward alcohol, and showed more overt behavioral signs of aversion during the taste test. This study is important in demonstrating that emetic aversion conditioning established conditioned aversion reactions to alcohol at posttreatment. The greater efficacy of emetic aversion treatment over shock aversion treatment was not surprising because this naturalistic experiment was conceptually similar to the cue-to-consequence experiment that demonstrated that taste-illness conditioning produced significantly stronger taste aversions than taste-shock conditioning (Garcia & Koelling, 1966).

It is worth noting that despite these positive reports there remains considerable debate about the effectiveness of emetic aversion conditioning as a treatment for alcoholism. Emetic aversion therapy is a treatment that requires highly trained personnel, sufficient time to implement, and a true commitment from the patient because the treatment is intense. The results from the well-controlled studies have shown emetic aversion therapy can be successful in producing short-term behavior change, and if the patient wants to curb his or her alcoholism, it is a valuable tool in helping him or her do so. Yet, like any form of taste-aversion learning, just as the organism can learn an aversion, it can also learn to extinguish that aversion. Therefore, if the patient is not highly motivated to change his or her alcoholic tendencies (or he or she is highly motivated to return to a life of alcoholism), he or she will be able to extinguish the alcohol aversions learned via emetic aversion conditioning after a number of months.

SUMMARY

Since John Garcia's initial research over 50 years ago, experimental investigations have revealed the robust and highly adaptive nature of taste-aversion learning with a wide range of species, stimuli, and conditions. An overview of this research reveals two overarching contributions from this work. First, as detailed in the Theory section, taste-aversion learning research has often identified phenomena that represent a challenge to existing learning theories or a departure from other types of classical conditioning. As a result, these findings have motivated considerable research and expanded the scope of theories of associative learning. Second, as detailed in the Applications section, the insight from the research has a number of practical benefits to humans. It will be interesting to see the developments produced by the next 50 years of taste-aversion research.

REFERENCES AND FURTHER READINGS

Baker, T. B., & Cannon, D. S. (1979). Taste aversion therapy with alcoholics: Techniques and evidence of a conditioned response. *Behaviour Research and Therapy, 17*, 229–242.

Barker, L. M., Best, M. R., & Domjan, M. (1977). *Learning mechanisms in food selection.* Waco, TX: Baylor University Press.

Barker, L. M., Smith, J. C., & Suarez, E. M. (1977). "Sickness" and the backward conditioning of taste aversions. In L. M. Barker, M. R. Best, & M. Domjan (Eds.), *Learning mechanisms in food selection*. Waco, TX: Baylor University Press.

Batsell, W. R., Jr., & Batson, J. D. (1999). Augmentation of taste conditioning by a preconditioned odor. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 374–388.

Batsell, W. R., Jr., & Best, M. R. (1992). Variations in the retention of taste aversions: Evidence for retrieval competition. *Animal Learning & Behavior, 20*, 146–159.

Batsell, W. R., Jr., & Best, M. R. (1993). One bottle too many? Method of testing determines the detection of overshadowing and retention of taste aversions. *Animal Learning & Behavior, 21*, 154–158.

Batsell, W. R., Jr., & Best, M. R. (1994). The role of US novelty in retention interval effects in single-element taste-aversion learning. *Animal Learning & Behavior, 22*, 332–340.

Batsell, W. R., Jr., & Brown, A. S. (1998). Human flavor-aversion learning: A comparison of traditional aversions and cognitive aversions. *Learning and Motivation, 29*, 383–396.

Batsell, W. R., Jr., Brown, A. S., Ansfield, M. E., & Paschall, G. Y. (2002). "You will eat all of that!" A retrospective analysis of forced consumption episodes. *Appetite, 38*, 211–219.

Batsell, W. R., Jr., & Paschall, G. Y. (in press). Mechanism of compound conditionong in flavor-aversion conditioning. In S. Reilly & T. Schachtman (Eds.), *Conditioned taste aversion: Behavioral and neural processes.* Oxford, UK: Oxford University Press.

Batsell, W. R., Jr., Paschall, G. Y., Gleason, D. I., & Batson, J. D. (2001). Taste preconditioning augments odor-aversion learning. *Journal of Experimental Psychology: Animal Behavior Processes, 27,* 30–47.

Batsell, W. R., Jr., & Pritchett, K. P. (1995). Retention of rotationally induced taste aversions. *Physiology & Behavior, 58*, 815–818.

Bernstein, I. L. (1978). Learned taste aversions in children receiving chemotherapy. *Science, 200*, 1302–1303.

Bernstein, I. L., Webster, M. M., & Bernstein, I. D. (1982). Food aversions in children receiving chemotherapy for cancer. *Cancer, 50*, 2961–2963.

Boakes, R. A., Tarrier, N., Barnes, B. W., & Tattersall, M. H. N. (1993). Prevalence of anticipatory nausea and other side-effects in cancer patients receiving chemotherapy. *European Journal of Cancer, 29A*, 866–870.

Bouton, M. E. (2006). *Learning and behavior: A contemporary synthesis*. Sunderland, MA: Sinauer Associates.

Braun, J. J., & McIntosh, H., Jr. (1973). Learned taste aversions induced by rotational stimulation. *Physiological Psychology, 1*, 301–304.

Broberg, D. J., & Bernstein, I. L. (1987). Candy as a scapegoat in the prevention of food aversions in children receiving chemotherapy. *Cancer, 60*, 2344–2347.

Cannon, D. S., Baker, T. B., & Wehl, C. K. (1981). Emetic and electric shock alcohol aversion therapy: Six- and twelve-month follow-up. *Journal of Consulting and Clinical Psychology, 49*, 360–368.

Domjan, M. (1980). Ingestional aversion learning: Unique and general processes. In J. S. Rosenblatt, R. A. Hinde, C. Beer, & M. C. Busnel (Eds.), *Advances in the study of behavior* (Vol. 11, pp. 275–336). New York: Academic Press.

Domjan, M., & Wilson, N. E. (1972). Specificity of cue to consequence in aversion learning in the rat. *Psychonomic Science, 26*, 143–145.

Dragoin, W., McCleary, G. E., & McCleary, P. (1971). A comparison of two methods of measuring conditioned taste aversions. *Behavior Research Methods & Instrumentation, 3*, 309–310.

Durlach, P. J., & Rescorla, R. A. (1980). Potentiation rather than overshadowing in flavor-aversion learning: An analysis in terms of within-compound associations. *Journal of Experimental Psychology: Animal Behavior Processes, 6*, 175–187.

Garb, J. L., & Stunkard, A. J. (1974). Taste aversions in man. *American Journal of Psychiatry, 131*, 1204–1207.

Garcia, J. (n.d.). *From nino to scientist: The trip of my life. Part I. Childhood: 1917–1932.* Unpublished manuscript.

Garcia, J., Ervin, F. R., & Koelling, R. A. (1966). Learning with prolonged delay of reinforcement. *Psychonomic Science, 5*, 121–122.

Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). A conditioned aversion towards saccharin resulting from exposure to gamma radiation. *Science, 122*, 157–158.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4*, 123–124.

Garcia, J., Lasiter, P. S., Bermudez-Rattoni, F., & Deems, D. A. (1985). A general theory of aversion learning. In N. S. Braveman & P. Bronstein (Eds.), *Experimental assessments and clinical applications of conditioned food aversions*, Annals of the New York Academy of Sciences (Vol. 443, pp. 8–21). New York: New York Academy of Sciences.

Grill, H. C., & Norgren, R. (1978). The taste reactivity test: I. Mimetic responses to gustatory stimuli in neurologically normal rats. *Brain Research, 143*, 263–279.

Grote, F. W., Jr., & Brown, R. T. (1971). Conditioned taste aversions: Two-stimulus tests are more sensitive than one-stimulus tests. *Behavior Research Methods & Instrumentation, 3*, 311–312.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church, (Eds.). *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Logue, A. W. (1979). Taste aversion and the generality of the laws of learning. *Psychological Bulletin, 86*, 276–296.

Logue, A. W. (1985). Conditioned food aversion learning in humans. In N. S. Braveman & P. Bronstein (Eds.), *Experimental assessments and clinical applications of* *conditioned food aversions, Annals of the New York Academy of Sciences* (Vol. 443, pp. 316–329). New York: New York Academy of Sciences.

Logue, A. W., Ophir, I., & Strauss, K. E. (1981). The acquisition of taste aversions in humans. *Behaviour Research & Therapy, 19*, 319–333.

LoLordo, V. M., & Droungas, A. (1989). Selective associations and adaptive specializations: Taste aversions and phobias. In S.B. Klein & R. Mowrer (Eds.), *Contemporary learning theories: Instrumental conditioning theory and the impact of biological constraints on learning* (pp. 145–179). Hillsdale, NJ: Erlbaum.

Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal Learning & Behavior, 4*, 186–192.

Morrow, G. R., & Dobkin, P. L. (1988). Anticipatory nausea and vomiting in cancer patients undergoing chemotherapy treatment: Prevalence, etiology and behavioral interventions. *Clinical Psychology Review, 8*, 517–556.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483–522.

Parker, L. A. (2003). Taste avoidance and taste aversion: Evidence for two different processes. *Learning & Behavior, 31*, 165–172.

Parker, L. A., Kwiatkowska, M., & Mechoulam, R. (2005). Delta-9-tetrahydrocannabinol and cannibidiol, but not ondansetron, interfere with conditioned retching reactions elicited by a lithium-paired context in *Suncusmurinus*: An animal model of anticipatory nausea and vomiting. *Physiology & Behavior*, 87, 66–71.

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.

Reilly, S., & Schachtman, T. R. (in press). *Conditioned taste aversion: Behavioral and neural processes*. Oxford, UK: Oxford University Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theories* (pp. 64–99). New York: Appleton-Century-Crofts.

Revusky, S., Parker, L. A., & Coombes, S. (1977). Flavor aversion learning: Extinction of the aversion to an interfering flavor after conditioning does not affect the aversion to the reference flavor. *Behavioral Biology, 19*, 503–508.

Riley, A., & Freeman, K. (2006). Conditioned taste aversions: An annotated bibliography. Retrieved 2006, from http://www.CTALearning.com

Rozin, P. (1986). One-trial acquired likes and dislikes in humans: Disgust as a US, food predominance, and negative learning predominance. *Learning and Motivation, 17*, 180–189.

Rozin, P., & Fallon, A. E. (1980). The psychological categorization of foods and non-foods: A preliminary taxonomy of food rejections. *Appetite, 1*, 193–201.

Rusiniak, K. W., Hankins, W. G., Garcia, J., & Brett, L. P. (1979). Flavor-illness aversions: Potentiation of odor by taste in rats. *Behavioral & Neural Biology, 25*, 1–17.

de Silva, P., & Rachman, S. (1987). Human food aversions: Nature and acquisition. *Behaviour Research & Therapy, 25*, 457–468.

Smith, J. C., & Roll, D. L. (1967). Trace conditioning with X-rays as an aversive stimulus. *Psychonomic Science, 9*, 11–12.

36

Operant Conditioning

Jessica G. Irons

James Madison University

William Buskist

Auburn University

If there is a single topic within psychology with which you are familiar, it is surely learning. After all, you have spent much of your life doing just that. As a child, you learned how to walk, how to talk, and how to get along (or not get along) with others. In school, you learned English, algebra, biology, geometry, and world history, among other things. And as an adult, you have continued to learn new things about yourself, your interests, and the world. With all that experience as a learner, you might think that you know everything there is to know about learning. But can you define what learning is? Can you describe how you have learned all the things that you have learned? Probably not—our experience as learners generally teaches us *what* we have learned but not *how* we have learned.

Understanding what learning is is complicated by the fact that there are so many ways that learning occurs. Nonetheless, psychologists generally define learning as a relatively permanent change in behavior and thinking based on experience. Learning cannot be observed directly; typically, psychologists infer that learning has occurred by noticing particular changes in behavior. However, behavior change is not by itself learning—it reflects only the possibility that learning has occurred. Thus, psychologists must be careful not to assume that changes in behavior always reflect learning. Consider, for example, how your behavior changes depending on whether you are wide awake or drowsy. Your arousal level may determine how well you perform on an examination, or the skill with which you execute a skilled activity such as playing a musical instrument or catching a baseball. Changes in behavior caused by arousal level, fatigue, ingestion of drugs, or emotional problems do not necessarily reflect learning.

Likewise, learning may occur without noticeable changes in behavior. In some cases, learning is not apparent, at least not right away, from changes in our observable behavior. In other cases, you may never have the opportunity to demonstrate what you have learned. For example, although you most likely learned how to change a flat tire in your high school driver's education course, your behavior will not reflect that learning unless you have the actual opportunity to change a flat tire. In still other cases, you may not be sufficiently motivated to demonstrate things that you have learned. Such might be the case when a teacher poses a question to the class: although you may know the answer, you do not say it because you just don't feel like raising your hand or you get nervous when speaking publicly.

As you can see, learning and behavior are closely but not perfectly related. Indeed, it is unusual for psychologists to speak of one without also speaking of the other. Psychologists study the relation between learning and behavior by manipulating the environment in which behavior occurs—that is, experience. In this chapter, we will focus on a particular form of learning—operant learning, which is better known as operant conditioning. We will outline the basic historical background of operant conditioning as well as the key terms, ideas, and theory relevant to understanding this very important type of learning.

BRIEF HISTORICAL OVERVIEW

Although operant learning (or conditioning) has existed since the beginning of human history, the study of the basic principles of operant conditioning is only about 100 years old. In fact, much of what we know today about operant learning resulted from the remarkable research of E. L. Thorndike and B. F. Skinner, who both designed clever experiments investigating how the consequences of one's actions influence the subsequent acquisition and maintenance of behavior.

E. L. Thorndike

Edward Lee Thorndike (1898–1911) enjoyed a prolific professional career, which included the study of children's mind-reading abilities, instinctive behavior in chickens, and how best to teach reading and spelling to schoolchildren (Thorne & Henley, 2001). However, Thorndike is perhaps most noted for his work in which he identified key principles of learning that turned out to be very similar to what we now call operant conditioning. Thorndike studied the abilities of cats to escape from puzzle boxes that he constructed from wood. Each puzzle box required a specific response or set of responses to be emitted in order for the cat to escape the box. During each trial, Thorndike watched each cat inside the box and carefully recorded what he observed. Over repeated trials, the cats learned to escape the box with increasing speed.

For Thorndike, the decreased latency between entry to the puzzle box and subsequent escape over repeated trials was indicative of improved performance, or learning. From this research, Thorndike developed his famous *Law of Effect*, which states that responses that produce satisfaction will be more likely to recur and thus be strengthened (Thorndike, 1911). In his own words:

> Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more formally connected with the situation, so that, when it recurs, they will be more likely to recur. (p. 244)

Likewise, a response followed by discomfort will be unlikely to recur and thus be weakened. Although Thorndike's work has been greatly advanced and improved in subsequent research, the law of effect remains an important one for understanding how operant learning occurs. As you read about Skinner's work with operant conditioning below, you will surely see Thorndike's influence on the field today.

B. F. Skinner

Like Thorndike, Burrhus Frederick Skinner (1904–1990) did his graduate work at Harvard University. Skinner championed the laboratory study of the law of effect and advanced its application to the study of the human condition and its attendant problems. He devised ingenious and rigorous laboratory methods for studying learning, invented unique laboratory apparatuses and methods for observing behavior, and created his own philosophy for interpreting it (Bolles, 1979).

Unlike many of his predecessors, Skinner believed that behavior should be examined for its own sake, and not only in terms of the underlying physiology or mental events that purportedly cause behavior. For Skinner, changes in behavior depend wholly on changes in environmental events. Skinner studied simple behavior under highly controlled laboratory conditions—for example, rats pressing a lever or pigeons pecking a plastic disk to access food. Skinner studied his subjects for long periods of time, as opposed to studying large numbers of animals for short periods of time. Skinner's brand of research, the scientific study of behavior, is known as *experimental analysis of behavior*. The application of knowledge gained from the experimental analysis of behavior to solving socially relevant issues is called *applied behavior analysis*, which is among the most vibrant fields of applied psychology today.

Skinner's (1938) *The Behavior of Organisms* presented one of his first articulations of what we now know as operant conditioning. By studying individual organisms in controlled laboratory apparatuses, called operant chambers, Skinner identified ways in which behavior may change as a result of manipulating environmental consequences. He also found that different schedules of the availability of consequences produced different effects on behavior. Finally, Skinner (1950) suggested that it is possible to study behavior and the processes of learning without a theory—a notion perhaps more radical than the notion of studying behavior for its own sake. The scope of operant conditioning has widened from studying the behaviors of Skinner's rats and pigeons to an array of complex human behaviors, such as reducing problem behaviors in children, making the workplace safer, and understanding and combating drug-use behavior.

In addition to his contributions to behaviorism and operant conditioning, Skinner's work led to the founding of two important psychological journals: the *Journal of Experimental Analysis of Behavior* (1958) and the *Journal of Applied Behavior Analysis* (1968). In addition to the well-known operant chamber, Skinner also invented several other novel apparatuses, including the cumulative recorder, the baby-tender, and the teaching machine. Finally, Skinner produced many notable writings, which remain influential today in the field of psychology. Among such writings are *The Behavior of Organisms*

(1938), *Schedules of Reinforcement* (with C. B. Ferster, 1957), and *Verbal Behavior* (1957). One book, *Walden Two*, describes a utopian society based on the application of operant principles to everyday living (Skinner, 1948).

OPERANT CONDITIONING

What Is an Operant?

The term *operant* refers simply to a behavior that is changed as a result of its consequences (Millenson & Leslie, 1979). The term *conditioning* refers to a specific type of learning. Thus, *operant conditioning* is learning produced as a result of behavioral consequences. More specifically, operant conditioning is behavior change that results from the *association* between or among responses and their consequences. The more frequently an operant occurs, the stronger it becomes (Skinner, 1938).

The Three-Term Contingency

Behavior does not occur in a vacuum; certain environmental events usually precede and follow behavior. Consider, for example, the practice of answering the telephone. The phone rings, you pick it up and say "hello" into the mouthpiece, and a person on the other end of the line begins to speak. How many times have you picked up the phone when it was not ringing and said "hello"? That would be a foolish thing to do, simply because no one is on the other end of the line with whom to speak. Certainly, your family, friends, or roommates would think you quite strange for this sort of behavior. We answer the phone (the behavioral event) because it rings (the preceding event) and because in the past someone has been at the other end of the line with whom to talk (the consequence, or following event). Skinner referred formally to the relationship among these three events as the *three-term contingency*. More specifically, the three-term contingency involves the following:

- The preceding event, or discriminative stimulus, which "sets" the occasion for responding because in the past that response has produced a certain consequence in the presence of that specific stimulus.
- The behavioral event or the response.
- The following event or consequence is dependent on the response. If the response does not occur, a consequence is not produced.

Thus, operant conditioning occurs in the presence of certain stimuli and is always followed by certain consequences. In the presence of a discriminative stimulus, a consequence will occur if and only if a response first occurs. The three-term contingency permits psychologists to understand any behavior that may occur in one context or situation (i.e., when a discriminative stimulus is present) but not in another (i.e., when the discriminative stimulus is not present). There are two basic types of consequences that may influence an operant: reinforcement and punishment.

TYPES OF BEHAVIORAL CONSEQUENCES

Reinforcement

Reinforcement occurs when an event results in the increased probability that the preceding behavior will recur on a future occasion. Or, put another way, reinforcement occurs if an event following a response increases subsequent responding. If a behavior increases following the addition of a stimulus (event or consequence), then that stimulus is called a *reinforcer*. A reinforcer is only available contingent upon a response—reinforcement occurs if and only if the organism emits a specific behavior. Although this principle may seem difficult to understand at first, the idea is one with which you are quite familiar. For example, if you are cold (a discriminative stimulus) and you put on a coat (the behavior), then you begin to feel warm (a reinforcing consequence). On future occasions when you feel cold, you will be more likely to put on a coat to help you feel warm because in the past you *learned* the association between the behavior of putting on your coat and warming up. In this case, the behavior of putting on a coat is *reinforcing*—warming yourself up is contingent upon putting on your coat. When a stimulus is added, resulting in an increase in behavior, we call that *positive reinforcement*. In operant conditioning, the term *positive* always refers to the addition of a consequential stimulus and does not imply a good or desirable consequence, although, to be sure, some added consequences are desirable.

Let's consider a second example of positive reinforcement. Suppose you tell a child to clean her room (the discriminative stimulus), and after she has cleaned the room (behavior), you give her a lollipop (a potential reinforcer). As a result, she is more likely to clean up her room in the future when asked because in the past she has received a lollipop for doing so. In this case, the lollipop served as a reinforcer for room-cleaning behavior.

In addition to adding a consequential stimulus to increase behavior, we can also subtract (or remove) a consequential stimulus to increase behavior. For example, suppose that you are wearing a coat and you become too warm. Of course, removing the coat will cool you off, which in this case is a pleasant condition for you. In this example, taking away the coat results in a consequence that increases the probability that you will remove your coat on future occasions when you feel hot while wearing it. The removal of a stimulus that results in an increased behavior is called *negative reinforcement*. The term *negative* in operant conditioning always refers to the subtraction or reduction of a consequential stimulus and does not imply a bad or undesirable consequence. Consider another example of negative reinforcement. Suppose you have a headache. You have certainly learned by now that when you have a headache, you can take an over-the-counter pain medication such as aspirin to quell or at least lessen the headache. Taking an aspirin (behavior) in the presence of a headache (the discriminative stimulus) results in its removal or reduction (the consequence), making it very likely that when you get a headache in the future, you will take aspirin (or some other pain reliever) again. The likelihood of taking aspirin again is *increased* as a result of *subtracting* the headache.

Keep in mind that it is all too easy to confuse positive and negative reinforcement. Here is a rule that will help you keep the two straight: Both positive and negative reinforcement increase the likelihood that a given behavior will recur. However, positive reinforcement involves the addition or presentation of a consequential stimulus, and negative reinforcement involves the subtraction, removal, or reduction (and sometimes the prevention) of a consequential stimulus.

Punishment

Punishment occurs when an event results in the *decreased* probability that the preceding behavior will occur on a future occasion: If an event following a response decreases subsequent responding, *punishment* has occurred. The stimulus that produces the decrease in behavior is called a *punisher,* and the decreased behavior is said to be *punished*. Punishment is also a result of a contingency—punishment will occur if and only if a specific behavior is emitted. Like reinforcement, punishment may also be either positive or negative. For example, suppose that you use more minutes than your cell phone plan allows for the month. As a result of your using extra minutes, the cellular company charges your account a fee for each additional minute. Additional charges decrease the likelihood that you will go over your minute limit in the future. In this case, using too many minutes is punished. The extra charges are contingent upon use of too many minutes and occur if and only if you use too many minutes. Notice that in this example, a consequential stimulus (the fee) is added to decrease behavior. A decrease in behavior as a result of adding a stimulus is called *positive punishment*.

Let's consider another example. Suppose a mother tells her child not to touch his grandmother's valuable collection of antique vases just prior to a visit to the grandmother's house. Despite the warning, the child picks up a vase and accidentally drops it, and the vase smashes into a thousand pieces all over the floor. Immediately, the mother scolds him. Chances are that the next time he visits his grandmother, he will refrain from touching the vase collection—his behavior of touching the vases will *decrease* as a result of the scolding (the addition of the scolding lessens the probability of vase touching in the future).

In addition to adding a consequential stimulus in order to decrease behavior, a consequential stimulus can also be subtracted to decrease behavior. For example, suppose that your cellular phone drops calls each time you walk into your kitchen. Dropped calls when you enter the kitchen decrease the likelihood that you will enter the kitchen while talking on your cell phone in the future. Notice that in this example, a consequential stimulus (losing a connection with the person with whom you were talking) was subtracted and produces a decrease in behavior (talking on your cell phone in the kitchen). A decrease in behavior as a result of subtracting a consequential stimulus is called *negative punishment*.

Consider another example. You are a high school student, and your parents have purchased tickets for you and a friend to see your favorite band in concert. Your father asks you to do your homework before bed one evening, and you refuse. As a result, your father takes away your tickets to the concert. On future occasions, you are more likely to obey your father's requests. In this case, your disobeying behavior is *decreased* by *subtracting* a consequential stimulus.

Note that psychologists always define reinforcement and punishment *functionally*—in terms of their effects on behavior. Consider a student whose teacher gives her a bonus point on an assignment for turning it in on time. If, on future occasions, she turns assignments in late, the bonus point did not reinforce or increase punctual behavior. Thus, even though the student received the bonus point, it had no effect on punctuality—the bonus point did not serve as a reinforcer. This example may seem counterintuitive, but it is illustrative of the importance of understanding how reinforcement and punishment are functionally defined. We can never know ahead of time if a stimulus or event is a reinforcer or punisher (or neither). We must always wait until after we have seen its effects on behavior before we can label it.

A second caveat is also in order: Psychologists restrict the use of the terms *reinforcement* and *punishment* to describe only environmental operations and their effects of the behavior of organisms. Behavior, not the organism, is said to be reinforced or punished.

The relationship between behavior and its consequences is often far more complex than can be described in this chapter. Indeed, some psychologists spend their entire careers trying to understand this complex relationship. In what follows, we will describe the most basic elements of operant conditioning by focusing primarily on positive reinforcement.

BASIC PRINCIPLES OF OPERANT CONDITIONING

Response Acquisition Through Shaping

How does a discriminative stimulus come to control behavior? How does an organism learn the relationship among discriminative stimuli, its behavior, and consequences? In short, how is behavior acquired? Most behavior is acquired gradually through the organism's interaction with reinforcing and punishing events in its environment. Skinner developed a technique for teaching new behavior to his nonhuman animal subjects called *shaping*. Shaping involves reinforcing any behavior that successfully approximates the desired response. Shaping maximizes the likelihood that the desired, or target, behavior will occur. As an organism's behavior more closely resembles the desired behavior, reinforcement is made available; however, if the organism's behavior differs from the target behavior, then reinforcement is not available.

Consider an example. Suppose you have a dog that you wish to "teach" to sit up on its hind legs. Although your dog does not spontaneously sit up on his hind legs, he will emit other behaviors such as sitting down on his hind legs, jumping up, or drawing his paws into the air in front of him. Each of these behaviors is not the target behavior, but each one is in some way similar to the target behavior or an approximation of the target behavior. Thus the process of shaping behavior is one that takes advantage of natural variability in behavior. Without variability, shaping could not occur, for behavior would never differ in any way to approximate a new response.

Let's get back to our example. You might present your dog with a tasty treat if he looks toward you when you say the words "sit up." Once that response is mastered, then you may present a treat only if the dog looks toward you and then sits. Once he is looking toward you and is sitting, you may reinforce a new, closer approximation of the target behavior—sitting up on his hind legs. If your dog fails to sit (but remains standing) on the next occasion on which you ask him to sit up, you would not present a treat because his behavior was not any more similar to the target behavior. However, if the dog looks at you while sitting and then makes any motion that resembles lifting its front legs off the floor, you present the treat. Gradually, each time you command "sit up," your dog will come closer and closer to the target behavior of sitting on his hind legs with both paws in front of his chest. The use of a reinforcer (e.g., a treat) shapes the dog's behavior of sitting up.

Let's walk through another example, this time featuring human behavior. Suppose you want your friend to be punctual each time she is to meet you. Being blunt, and getting upset that she is late all the time, is not always the most effective way to change behavior while also trying to maintain a relationship. In this case, shaping represents an effective alternative method for changing her behavior. Suppose that your friend is typically about 15 minutes late to meet you and on one occasion she is only 10 minutes late. On this occasion, offer a potential reinforcer—perhaps simply by telling her how much you appreciate that she was almost on time. The next time you meet your friend, offer reinforcement if she is less than 10 minutes late, and so on until your friend is on time. Who knows—maybe she will even be early someday!

As you may have guessed, the trick to shaping is to identify an effective reinforcer. Recall that not all potential reinforcers turn out to be reinforcing—only those that increase behavior will be true reinforcers. If you buy your friend a soda when she is on time, and she does not continue to be on time on future occasions, the soda did not serve as a reinforcer. She may prefer verbal praise or a smile or a hug. Identifying which potential reinforcers may actually shape behavior is generally a trial-and-error process—but the better you know a person, the more you will know about the things she likes or doesn't like.

Shaping is a useful procedure both in the laboratory and in the real world. Skinner shaped rats to press levers and pigeons to peck colored disks to gain access to food in an operant chamber. Parents use shaping to teach their children to lace their shoes, get dressed, use language, catch a ball, ride a bicycle, and many other behaviors.

One advantage of the shaping process is that behavior can change in response to the environment without the need to use verbal instructions telling the organism what to do. In effect, reinforcement "tells" the organism how to behave. Generally speaking, reinforcement and its absence are used in shaping—punishment is not.

Response Discrimination

In many instances, only one aspect of behavior produces reinforcement in a particular circumstance, and all others do not. This process of an organism learning which aspects of a response produce reinforcement is called *response discrimination*. It develops as the organism learns that only certain features of a response such as its speed, accuracy, and frequency, are reinforced in any particular context. For example, when teaching your dog to come to you when you utter the command "come," you might give it a treat if and only if it runs toward you. You do not provide it a treat for walking or crawling toward you.

Response Generalization

In some cases, different aspects of the same behavior may produce reinforcement in the presence of the same context. This process is called *response generalization*. In contrast to response discrimination, response generalization occurs when more than one feature of behavior will produce the same reinforcer. For example, suppose that you wish to teach your dog to come toward you when you say the word "come." Each time it does, you give it a treat—but it doesn't matter how the dog approaches you. It could walk or run or even dance its way toward you—it still approaches you when you command it to come.

Extinction

Any procedure in which a reinforcer is no longer presented following a response is called *extinction*. Behavior that is no longer reinforced decreases in frequency until it stops occurring altogether. Such behavior is said to extinguish because it is no longer emitted. Suppose you put money into a vending machine to purchase a soft drink and nothing comes out. Will you keep putting money into the machine? Not likely. One interesting consequence of extinction is that it initially produces response variability. When the soft drink machine is working properly, your behavior is simple. You drop several coins into the machine, push a particular button, and out comes your ice-cold soft drink. But what happens when you put your money into the machine, push the button, and nothing happens? Now your behavior changes: You may press the button harder or start pushing down on the coin return lever. If these behaviors fail to either produce your soft drink or return your money, you may start pushing or hitting the machine, or perhaps even yell at it, as if it could understand human language and feel guilty for not giving you what you wanted! Eventually, you stop engaging in all these behaviors and walk away from the mechanical bandit. Thus, extinction usually produces an increase, then a decrease, in behavior.

Spontaneous Recovery

Spontaneous recovery occurs when a previously extinct behavior reemerges, even without further association of a reinforcing or punishing consequence with the target behavior. For example, suppose a few days have passed since you lost your money in the vending machine. You decide to give it another try, on the spur of the moment, even though it "robbed" you last time. Such behavior following extinction is called *spontaneous recovery*. Interestingly, if the machine is now working properly, and you receive your soft drink after you pay for it and press the button, your "vending machine" behavior is actually strengthened.

Continuous and Intermittent Reinforcement

Broadly speaking, reinforcement of a response can occur continuously or intermittently. Behavior that is reinforced each and every time it occurs is said to be reinforced continuously. Behavior that is only occasionally reinforced is said to be reinforced intermittently. Not surprisingly, these two reinforcement processes have different effects on behavior. Skinner (1956) first demonstrated this finding when he was running low on food supplies for his rats and he attempted to conserve food by only occasionally reinforcing his rats' responding instead of after each response. He discovered that this procedure had two effects. First, of course, it increased the longevity of his food supply. Second, and more important, compared to continuous reinforcement, intermittent reinforcement produced large amounts of responding in his rats. He also later found that during extinction conditions, rats whose responding was intermittently reinforced produced a greater number of responses during extinction conditions than did rats whose behavior had been continuously reinforced. Thus, training under intermittent reinforcement made his rats' behavior more resistant to extinction—or more durable than training under continuous reinforcement.

REINFORCEMENT PARAMETERS

Several important variables influence the efficacy of potential reinforcers for maintaining learned behavior. For example, reinforcement history or a history of experience with specific reinforcers is a powerful determinant of behavior. The more often you encounter the same association of a specific behavior (e.g., putting on a coat) and its specific consequence (e.g., feeling warm), the stronger your reinforcement history is likely to be. A strong reinforcement history with any particular reinforcing stimulus increases the likelihood that a behavior will occur on future occasions. Other factors that influence how reinforcement influences behavior include reinforcer magnitude, response requirements to gain access to reinforcers, and the schedule of availability of potential reinforcers.

Reinforcer Magnitude

Reinforcer magnitude refers to size or amount (e.g., frequency or duration) of a particular reinforcing stimulus or event. In general, the greater the magnitude of a reinforcing stimulus, the more likely it will influence behavior upon its presentation. This concept makes sense when you stop and think about the wages people are willing to work for in their jobs. Most people would rather make more money than less money—so a job that pays $20 per hour is more attractive than a job that pays only minimum wage.

Response Requirements

Response requirement refers to the amount of behavior (e.g., the number of responses, intensity of responses) that must be emitted to obtain a reinforcer. In general, the lower response requirement required to produce a reinforcer, the more likely an organism is to engage in the behavior necessary to acquire it. For example, consider a study in which adult men pulled a lever to gain access to alcohol (Van Etten, Higgins, & Bickel, 1995). Participants could gain access to alcohol after lever pulling from a low response requirement of 100 times to a high requirement of 1,600 times. Alcohol served only as a reinforcer for lower response requirements (e.g., less than 800 lever pulls) but not for high response requirements. Thus, some response requirements (lower requirements) do not affect the efficacy of reinforcers and other response requirements (higher requirements) render potential reinforcers ineffective for changing behavior.

Schedules of Reinforcement

Your own experience has no doubt taught you that there is not always a one-to-one correlation between your behavior and its consequences. Sometimes your behavior is reinforced by its consequences and sometimes it is not. Dialing the telephone sometimes results in a busy signal, sometimes in no answer, and sometimes in the expected "hello." Sometimes you get away with exceeding the posted speed limit on the freeway and sometimes you do not. Depending on the behavior, reinforcement and punishment often vary in frequency over time. To study the effects of such variation, psychologists have devised schedules of reinforcement that specify the conditions that must be satisfied before an organism's responding is reinforced (Ferster & Skinner, 1957). These conditions are usually based on the passage of time, the number of responses emitted by the organism, or some combination of both.

Skinner's discovery of intermittent reinforcement in maintaining behavior stimulated an enormous amount of research on the effects of different kinds of reinforcement schedules on human and animal behavior. Indeed, many pages of the *Journal of the Experimental Analysis of Behavior* are filled with detailed information about the precise ways in which these schedules influence behavior. Four kinds of schedules—the so-called simple schedules—have received the most research attention. Skinner and others created these schedules by combining fixed or variable delivery of reinforcement with response-based or time-based delivery of reinforcement. Hence, their names: fixed-ratio, variable-ratio, fixed-interval, and variable-interval. That these four schedules have different effects on behavior is evidenced by the fact that the same animal will produce different patterns of responding under each schedule. Let's look at each of these schedules in a bit more detail.

Fixed-Ratio (FR) Schedules

Fixed-ratio or FR schedules require that reinforcement be delivered only after the organism emits a fixed number of responses. Thus, in an FR 10 schedule, every 10th response is reinforced. As you might expect, FR schedules produce high rates of responding because the rate of reinforcement is directly correlated with the amount of responding: The greater the number of responses the organism emits, the greater the number of reinforcements it receives. Organisms responding under FR schedules typically pause briefly after receiving reinforcement but then respond rapidly until they receive the next reinforcer.

Variable-Ratio (VR) Schedules

The response requirement for reinforcement under variable-ratio or VR schedules varies from one reinforcer delivery to the next. For example, in a VR 100 schedule, the reinforcer may be delivered after the 50th response one time, after the 150th response the next time, and after the 100th response the following time. Thus, in a VR 100 schedule, the average number of responses to produce a reinforcer is 100. VR schedules produce higher rates of behavior than do FR schedules—and that behavior is highly resistant to extinction.

Fixed-Interval (FI) Schedules

With fixed-interval or FI schedules, only the first response that occurs after a fixed amount of time has elapsed is reinforced. Thus, in an FI 20-second schedule, the first response after each 20-second interval has elapsed since the previous reinforcer delivery (or the start of the experimental session) is reinforced. All other responses have no effect on the delivery of reinforcement. For many animals, FI schedules produce a "scalloped" pattern of responding: immediately after reinforcement there is no responding—there is a post-reinforcement pause in responding. As time passes, though, responding increases gradually to the point of the next reinforcer delivery.

Variable-Interval (VI) Schedules

Like VR schedules, the response requirement for reinforcement under variable-interval or VI schedules varies from one reinforcer delivery to the next. Unlike VR schedules, though, the criteria for reinforcement is more time-based than response-based. For instance, in a VI one-minute schedule, responses are reinforced, on average, every one minute since the last reinforcer delivery. Some reinforcers are delivered after short intervals and some are delivered after longer intervals, producing a moderate steady rate of responding. In all cases, though, the first response that the organism emits after the interval times out produces a reinforcer. Behavior learned under VI schedules, like behavior learned under VR schedules, is highly resistant to extinction.

Although it is tempting to look for these sorts of reinforcement schedules in everyday life, they are more likely to be found within the confines of the controlled laboratory environment. Nature, it seems, does not parcel out its reinforcers in neat little packages that arrive according to one specific schedule (Crossman, 1983). For example, although most people receive a paycheck after fixed intervals of time, they generally do not pause before continuing to work—if they did, they would get fired! Too many other factors including peer pressure, natural enjoyment of work, and fear of being fired make it unlikely that a postreinforcement pause will occur following getting paid. Nonetheless, schedules of reinforcement have been tremendously useful tools to psychologists in understanding how specific environmental events influence behavior (Mazur, 2002).

Other Reinforcement Schedules

In addition to the simple schedules, there are also many more complex schedules of reinforcement. For example, in *concurrent schedules,* reinforcement is available under two (or more) schedules operating independently of each other. The most common concurrent schedule is the concurrent VI VI schedule. For example, imagine a pigeon in an operant chamber with two colored disks that may be pecked to produce access to grain. The left disk operates on a VI

20-second schedule, while the right disk operates on a VI 60-second schedule. The pigeon may respond by pecking on either disk for grain because both VI schedules are available *concurrently*. Because concurrent schedules generally operate independently of one another, responding on one schedule does not influence the availability of reinforcement on the other schedule. Suppose the pigeon in our example responded on each disk—pecking the left disk would not change the availability of grain on the right disk.

Concurrent reinforcement schedules produce a fascinating behavioral phenomenon called *matching*. In our example of the pigeon responding under the concurrent VI 20, VI 60 schedule, the pigeon is highly likely to respond three times more on the VI 20 schedule relative to the VI 60 schedule because the ratio of reinforcement is three to one across the two schedules (for every minute that passes under these two schedules, reinforcement on the VI 20 schedule is three times more likely than reinforcement on the VI 60 schedule). Matching has received intense empirical scrutiny since it was first discovered in the early 1960s by Richard Herrnstein (1961, 1970; for a detailed review, see Davison & McCarthy, 1988)

Up to this point, our discussion of concurrent schedules suggests that the reinforcement available on each schedule is always the same. It does not have to be: Each independent schedule may have a different type of reinforcement associated with it. For example, a pigeon may peck the right disk to gain access to grain and peck the left disk to gain access to water. The responses to each disk are the same (pecks on the right and pecks on the left), but in this case, the reinforcers are different.

Other types of complex reinforcement schedules include multiple schedules and mixed schedules. *Multiple schedules* of reinforcement are similar to concurrent schedules except that the schedules of reinforcement are in effect one at a time, and a specific stimulus indicates when each is available. For example, a pigeon behaving under a multiple schedule for access to grain might see a green light when an FI five-second schedule is in effect and a red light when a VR 20 schedule is in effect. The lights (i.e., red and green) indicate which schedule is in effect. Thus the lights serve as discriminative stimuli. *Mixed schedules* of reinforcement are functionally the same as multiple schedules, except that there is no discriminative stimulus indicating which schedule is in effect.

CONDITIONED REINFORCEMENT

To this point, we have discussed reinforcement only in terms of primary reinforcers—biologically relevant stimuli such as food, water, and painful events. However, as you know from personal experience, behavior can also be maintained with a wider variety of stimuli—money, a smile, a hug, kind words, a pat on the back, and prizes and awards. These stimuli acquire their reinforcing properties through their association with primary reinforcers and are called *conditioned reinforcers*. Because it can be exchanged for so many different kinds of primary reinforcers in our society, money is by far the most common conditioned reinforcer among humans. That money is surely a conditioned reinforcer can be substantiated by asking yourself a simple question: would you continue to work if money could not be exchanged for food, water, shelter, medical services, and other necessities of life?

Conditioned reinforcement is the principle on which token economies are based. A token economy is a system used in some mental institutions and other applied settings, such as schools, to engender desired behavior. For example, in a mental institution, a resident might be given small plastic tokens for emitting desired behavior such as that involved with personal hygiene and social interaction. These tokens are later exchangeable for different goods and services. In the classroom, students who engage in appropriate behavior may earn tokens that are exchangeable for snack items and school supplies.

STIMULUS CONTROL

You may recall from our earlier discussion of response discrimination and response generalization that sometimes only a particular type of response will produce reinforcement in the presence of a discriminative stimulus, and sometimes variations of this response will produce reinforcement in the presence of the discriminative stimulus. In this section, we consider *stimulus* discrimination and *stimulus* generalization, or what is better known as *stimulus control*. This term implies that environmental stimuli exert strong control over behavior because of the consequences that are associated with them.

Stimulus Discrimination

Basically, stimulus discrimination is the process of learning about which stimuli indicate the potential for reinforcement, and which do not. Another way to think about stimulus discrimination is that behavior changes depending on the context. Suppose a pigeon receives grain only for pecking a green disk and not for pecking a red disk. The pigeon quickly learns to peck at the green disk and not to peck at the red disk. Thus, the pigeon learns to *discriminate* between the two different colored disks. Consider another example. Suppose that a smile on your mother's face indicates that she is likely to give you money if you do the dishes. The money is available if, and only if, the smile is present and you do the dishes. If you do the dishes and the smile is not present, then she is not likely to offer you any money for doing the dishes. If you wash the dishes only when your mother is in a good mood (i.e., smiling), you have discriminated among her facial expressions to determine which ones indicate the likelihood of monetary reinforcement and which do not. Stimulus discrimination occurs when an organism behaves (i.e., does the dishes) in the presence of a particular stimulus (e.g., a smile) but not in the presence of other stimuli (e.g., a frown or a scowl). Stimulus discrimination results in specific behaviors in response to the presence of a specific discriminative stimulus.

Let's consider one last example of stimulus discrimination. In order for your dog to learn that a treat is available for sitting up only when you utter the words "sit up," you must give your dog the treat only when you say those words—they constitute the discriminative stimulus. If the dog lies down, rolls over, barks, puts its nose in your lap, or does anything else, you would not give your dog a treat. After repeated trials of sitting up following your command to "sit up" and following such behavior (and only such behavior) with a treat, the dog will learn to sit up only when you give this command.

Stimulus Generalization

Stimulus generalization refers to the process of determining which stimuli, in addition to a specific discriminative stimulus, may also function as a cue for the availability of reinforcement. The closer stimuli resemble the discriminative stimulus on which they were originally trained, the more likely organisms are to respond to those stimuli. Another way to think about this process is that learning in one situation influences how you behave in similar situations. Consider again a pigeon in an operant chamber. We have trained it to peck a green disk to access grain. What would happen if we presented a greenish-blue disk to the bird? Perhaps it would peck at it because the disk resembles the green disk we originally trained it to peck. What if we present the pigeon with a blue disk? If the pigeon pecks the green-blue disk and the blue disk, then we can safely say that stimulus generalization has occurred.

Similarly, we often observe (sometimes amusing) confusion when we try to teach children to learn to name different things such as animals. For example, we might be trying to teach a child to say the word "dog" each time it sees a dog. In a short while, the child will learn to say the word "dog" each and every time she sees a dog. But then, suppose we are out in the country riding in our car when the child spots a cow and spontaneously calls it a "dog." In this case, the child has generalized the word "dog" to apply to another four-legged creature—a cow. (Now, our task as a teacher becomes one of teaching the child not to generalize, but to discriminate, so that she learns to say "dog" each time she sees a dog, and to say "cow" each time she sees a cow.)

Stimulus discrimination and generalization don't just occur with the family pet and with learning to name animals. Such behavioral processes are ubiquitous in human society. For example,

- we cross the street when the traffic light is red and vehicles have stopped on the street we wish to cross, but not when the traffic light is green and cars are oncoming (response discrimination);
- we sit quietly in our seats during a lecture, church services, and theatrical presentations (response generalization);
- we raise our hands before speaking in class, but not while partying with friends (response discrimination);
- we say "thank you" to people who have done something nice for us (response generalization).

Let's return to our pigeon example one more time. Suppose that we train the pigeon to peck at the green disk—pecks at this disk occasionally produce grain (let's say we're using a VI 60-second schedule). Suppose now that we begin to change the color of the disk, so that the basic shades of the color spectrum appear over several sessions in the operant chamber. What do you think the pigeon will do? How will it behave? Which disk colors will generate the most responding from our winged friend? Well, if we were to plot the frequency of pecking as it corresponds to different shades of the disk, we would discover that the pigeon will peck most frequently when the disk is green and next most often at the colors that are similar to green—but that it would peck much less often, if at all, when the color of the disk is very different from green—say, a bright orange or red. This plot of the frequency of behavior generated by these different colored stimuli is called a *generalization gradient*. More specifically, most responding will likely occur at the point of the original discriminative stimulus (i.e., green). As the shades of color vary away from green to yellow or blue, the bird will likely continue to peck at a moderate rate because yellow and blue are most similar to green on the color spectrum. Thus, as the shades of color vary further and further from green—for example, orange and red— pecking will likely be very low frequency because these colors are very different from green.

Taken together, the phenomena of stimulus discrimination and stimulus generalization offer insight into some of the ways in which behavior comes under the control of environmental stimuli. Up to this point, we have discussed behavior control and behavior change largely as a function of positive reinforcement. In our final section, we outline ways in which negative reinforcement influences behavior.

AVERSIVE CONTROL OF BEHAVIOR

An aversive stimulus is one that an organism will behave to avoid or escape or reduce (Fantino, 1979). You will recall that negative reinforcement is any stimulus or event that, when removed, reduced, or prevented following a response, increases the frequency of that response over time. You also will recall that punishers decrease behaviors upon which they are contingent. Typically, punishers are aversive stimuli. Aversive control of behavior is the use of aversive stimuli to control behavior—the idea is that organisms, particularly people, will act to avoid or reduce the likelihood that they receive or encounter aversive stimuli. To be sure, such control of behavior seems to pervade every nook and cranny of our society. From the spankings given to misbehaving toddlers and the fines given to speeding motorists, to assigning poor grades to subpar students and imprisoning felons for their criminal deeds, our society uses aversive stimuli to attempt to control the behavior of its citizens.

Aversive control is prominent for two reasons. First, it can be highly effective in promoting behavior change, producing nearly immediate results. A person given a hefty fine for running a stop sign is likely, at least for a short while, to heed the sign's instruction in the future.

Second, in many cases, society has little control over the positive reinforcers that shape and maintain individual behavior. However, society has a great deal of control over aversive stimuli that can be used to punish misconduct. For example, although society has no control *per se* over the kinds of stimuli that positively reinforce crime (money and power), it does control the stimuli for punishing such behavior (fines, community service, jail time). Although your driver's education teacher probably complimented you when you obeyed traffic signs, when was the last time a police officer (or, for that matter, anyone else) rewarded you for such obedience? Your answer, of course, is likely to be "never" because in this case the police do not possess stimuli that can be used to reward your good driving habits—but be assured, they surely possess the stimuli that can be used to punish your actions. The theory of good driving behavior is that drivers will act to avoid getting pulled over, receiving a fine, or being sentenced to jail.

Although aversive control of behavior is effective in reducing and in some cases eliminating undesirable behavior, it can also produce several extremely negative side effects:

- Unrestrained use of physical force by the person doing the punishing may cause serious bodily injury to the person being punished (e.g., child abuse).
- Negative associations become attached to the people meting out punishers; punishment often induces fear, hostility, and other unwanted emotions in the people who are being punished. It may even result in retaliation against the person doing the punishing.
- Through punishment, the person being punished learns only which response NOT to emit—punishment does not involve teaching the person being punished how to emit the correct or desirable behavior.

Thus, although aversive control of behavior is a popular means for governing individual behavior, it has several drawbacks. And, although the effects of aversive control can be reinforcing for persons doing the punishment, the persons on the receiving end often receive very little personal benefit. This effect is one reason why Skinner advocated that society use positive reinforcement rather than negative reinforcement to govern its citizenry.

SUMMARY

Much of our behavior, as well as the behavior of many other animals with whom we share this planet, is governed by operant conditioning. Skinner's three-term contingency explains behavior acquisition, change, and maintenance through antecedents (discriminative stimuli), responses, and consequences. There are two basic types of consequences for behaving: reinforcement and punishment. Reinforcement occurs when some event results in the increased probability that the preceding behavior will occur on a future occasion and can be either positive (adding a stimulus) or negative (removing a stimulus). Punishment occurs when some event results in the decreased probability that a behavior will occur on a future occasion and can also be positive (adding a stimulus) or negative (removing a stimulus). Some stimuli, called primary reinforcers such as food and water are natural; other stimuli, called conditioned reinforcers such as money gain their reinforcing power by becoming closely associated with primary reinforcers.

The consequences of behavior play an integral role in the modification of responding. Shaping is a key procedure used to change behavior—it works through the application of reinforcement to successive approximations of a target behavior. Schedules of reinforcement also influence the acquisition and maintenance of behavior. Different kinds of reinforcement schedules produce different patterns of responding. The stimuli associated with reinforcement—discriminative stimuli—exert powerful control over behavior. Aversive control of behavior has been shown to be effective in controlling behavior, but it has several drawbacks that limit its overall ability to control behavior.

REFERENCES AND FURTHER READINGS

Bolles, R. C. (1979). *Learning theory.* New York: Holt, Rinehart & Winston.

Crossman, E. K. (1983). Las Vegas knows better. *The Behavior Analyst, 6,* 109–110.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Hillsdale, NJ: Erlbaum.

Fantino, E. (1979). Aversive control. In J. A. Nevin & G. S. Reynolds (Eds.), *The study of behavior: Learning, motivation, emotion, and instinct* (pp. 238–279). Glenview, IL: Scott, Foresman and Company.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement.* New York: Appleton-Century-Crofts.

Herrnstein, R. J. (1961). Relative and absolute strength of responses as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4,* 267–272.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13,* 243–266.

Mazur, J. E. (2002). *Learning and behavior* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Millenson, J. R., & Leslie, J. C. (1979). *Principles of behavioral analysis* (2nd ed.). New York: Macmillan.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1948). *Walden two.* New York: Macmillan.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review, 57*, 193–216.

Skinner, B. F. (1956). A case history in the scientific method. *American Psychologist, 11*, 221–233.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* New York: Macmillan.

Thorne, M., & Henley, T. B. (2001). *Connections in the history and systems of psychology* (2nd ed.). Boston: Houghton Mifflin.

Van Etten, M. L., Higgins, S. T., & Bickel, W. K. (1995). Effects of response cost and unit dose on alcohol self-administration in moderate drinkers. *Behavioural Pharmacology*, *6*, 754–758.

37

Recent Trends in Operant Conditioning

Harold L. Miller, Jr., and E. Benjamin H. Heuston
Brigham Young University

Our starting point for this chapter was a frequency analysis of the key words attached to articles published during the 10-year period from 1996–2006 in the *Journal of the Experimental Analysis of Behavior*, a standard source of literature on operant conditioning. The topics we take up here are those our analysis revealed as more readily definitive, though we do not present them in order of relative frequency. Taken together, we consider them a reasonable roadmap for navigating the present landscape of theory and research in operant conditioning. We will turn to the topics after a brief look at the basic elements of the foundational theory of operant conditioning and general methods for its study.

BASIC THEORY AND METHODS

The word *theory* rarely appeared among the key words. This absence may reflect a lingering ambivalence about the role of theory in operant conditioning, an ambivalence traceable to the work of the individual whose name appeared most frequently in our analysis—B. F. Skinner. His illustrious career spanned more than 60 years and saw him in several distinct roles as researcher, innovator, author, teacher, philosopher, social commentator, and, for some, visionary. He coined the term *operant conditioning*. No name is linked more canonically to operant conditioning than his. His first book, *The Behavior of Organisms*, was published in 1938 and remains unrivaled as a source of theory about operant conditioning (see the special issue of *Journal of the Experimental Analysis of Behavior* published in 1988 [Vol. 50, No. 2] that marked the 50th anniversary of *The Behavior of Organisms*). The book reports his ingenious research efforts to disentangle the operant form of behavior from the reflexive forms that others had previously identified and that constituted psychology's preoccupation at the time. The research was conducted with rats. Later he introduced pigeons and eventually humans as research subjects. Indeed, these three categories of research subject were the most frequent entries in our analysis of key words, *pigeons* and *rats* appearing much more frequently than *humans*. The other most frequently appearing terms were the response forms that Skinner introduced for his animal subjects—*key peck*s and *lever presses*, respectively.

Though *The Behavior of Organisms* may be viewed as laying out a theory of operant conditioning, Skinner's (1953) other most influential book, *Science and Human Behavior*, is skeptical about the role of theory in the science of behavior (see the special issue of the *Journal of the Experimental Analysis of Behavior* published in 2003 [Vol. 80, No. 3] that marked the 50th anniversary of *Science and Human Behavior*). Or at least he was skeptical about a certain type of theory—namely, one that uses one set of terms to postulate about the causes of behavior and a different set of terms to describe the actual behavior of interest. Skinner was particularly harsh in rejecting mentalistic theories of behavior—theories that referred to mind and thought, to will and agency, and so on. He considered theories with these fictional entities at their heart as frivolous, or worse, as unscientific impediments to a real science of behavior, a position he championed to the end of his life (see Skinner, 1990).

To make his own theoretical position clear, Skinner articulated the well-known *three-term contingency of reinforcement*. In addition to behavior (measured as discrete responses belonging to an operant class and often expressed in terms of the rate of responding—that is, the number of responses over an interval of time), he included the antecedents of responses (*discriminative stimuli*, the discrete stimuli that "set the occasion" for responses [in his phrase]) and the consequences of those responses (specifically, the stimuli that responses produce or that otherwise follow responses), which he termed *reinforcers* (and which are also often measured in terms of rate). For Skinner, stimuli acquire their discriminative function through conjunction with the relations between responses and reinforcers, as when reinforcers increase the rate of responding, and the absence of reinforcers decreases it. The science of behavior will proceed best by analyzing behavior at the level of discriminative stimuli, responses, and reinforcers and in terms of the contingent and noncontingent relations between them. In other words, Skinner advocated a functional analysis of behavior, an analysis of how it functions amid discriminative stimuli and reinforcers. His invention of *schedules of reinforcement* constituted an explicit means of analysis—methods for instantiating stimuli, responses, and consequences and for understanding their relations. A schedule is a collection of procedural steps whereby discriminative stimuli, responses, and reinforcers are presented in the course of an experimental session (see Ferster & Skinner, 1957, for an encyclopedic look at schedules and their effects on pigeons, rats, and humans).

Skinner's commanding importance to operant conditioning is reflected in the regularity with which *The Behavior of Organisms* and *Science and Human Behavior* continue to be cited as the jumping-off place for research and theory. As a recent example, Catania (2005b; see also Killeen, 1988) returned to Skinner's (1938) original concept of the *reflex reserve*, which he later renamed the *operant reserve* but eventually abandoned. According to Skinner, the operant reserve was a hypothetical store of responses that operated hydraulically. It rose when responses were reinforced and fell when they were emitted but not reinforced. The concept ran into trouble when faced with the so-called *partial reinforcement effect*, the demonstration that responses reinforced only intermittently (exemplified by schedules of reinforcement in which only some, not all, responses were followed by reinforcers) were more resistant to extinction. According to Catania, it is possible to save the concept of a reserve by extending the influence of a specific reinforcer not just to the reinforced response but also to the other, nonreinforced responses that preceded it. In this way, each response adds to the reserve, though not to the same degree, as well as depletes it. At issue is the extent to which nonreinforced responses increase the reserve. Catania considered various ways of expressing that contribution as a function of the time that elapsed between a response and the next occurrence of a reinforcer—the *delay-of-reinforcement gradient*. He also applied this view to several different types of reinforcement schedule, using computer simulations to suggest the type of performance the reserve would generate under each type.

Catania's (2005b) article also exemplifies the growing use of mathematics to do more than merely summarize research results. He used mathematical functions to describe hypothetical processes that are responsible for the relations between responses and reinforcers. Mazur (2006) recently documented the increased use of mathematical models with a frequency analysis of the contents of *Journal of the Experimental Analysis of Behavior*. From 1960 through 2000, the percentage of articles that included mathematical equations increased approximately tenfold, from 3 to 30 percent. Reflecting on this growth, Mazur cited several benefits of the reliance on mathematical models in behavior analysis. Specifically, they lend improved precision over purely verbal descriptions of data and processes, thereby sharpening the differences in predictions between competing accounts of the behavior in question. In doing so they may call attention to factors that were overlooked and to the insufficiency of competing accounts, thereby stimulating their improvement. Mathematical models may also provide an integrative role, drawing together data from disparate sources in the interest of identifying a core set of theoretical concepts by which to examine and account for a growing diversity of behaviors. In addition, Mazur pointed to use of mathematical models for improved communication between researchers and between researchers and practitioners, including behavior therapists.

To illustrate the role of mathematical models in deciding between competing theoretical views, Mazur (2006) focused on the use of conditioned reinforcers in higher-order schedules of reinforcement known as *concurrent chains*. Conditioned reinforcers may result from pairing originally neutral stimuli with biologically significant reinforcers such as food or water and are widely available in human experience. In concurrent-chains schedules, conditioned reinforcers may be colored lights that follow responses to the plastic discs that pigeons peck in order to produce food. Mazur presented three different models of behavior in concurrent-chains schedules. For each, he stated the prediction the model makes as a function of the delay to reinforcement (the same variable that Catania used in the operant reserve study)—in this case, the delay between the appearance of the conditioned reinforcer and the delivery of food. Future experiments designed to test the predictions may reveal the relative superiority of one model over the others and lead to subsequent revision of the latter or their abandonment in favor of the former.

Mazur's reference to mathematical models that integrate diverse phenomena is exemplified by Killeen's Mathematical Principles of Reinforcement (MPR), a theory of reinforcement schedules that combines principles of motivation, constraint, and association, each specified mathematically. It begins simply enough with reference to the behavioral arousal that reinforcers elicit and to the relation among arousal, rate of reinforcement, and rate of response. It then considers the way in which the minimal amount of time required to emit a response may constrain the rates of response and reinforcement when schedules of reinforcement are imposed. Finally, it makes room for a now-familiar variable: the delay between a response and a reinforcer (i.e., the extent to which they may be associated). In mathematical terms, MPR is expressed as follows:

$$b = \frac{cr}{\delta + 1/a}$$

where b stands for response rate and r for reinforcement rate, a is a measure of arousal level, δ represents the average amount of time required to emit a response, and c is a measure of the association between response and reinforcement specific to the reinforcement schedule that is in place. Mazur cited the extension of MPR to a wide range of behavioral variables, including the effects of brain lesions and drugs.

In addition, Mazur (2006) cited the relevance of MPR to what has become the overarching issue in operant conditioning, namely, the level at which behavior analysis properly takes place. In *The Behavior of Organisms,* Skinner (1938) gave meticulous attention to specific properties of the rat's response itself, including its form (topography), duration, and force, only to conclude eventually that rate of responding was the primary datum of operant conditioning. This distinction established the counterpoise between studying the effects of reinforcement at the momentary level at which responses occur (sometimes referred to as the *local* or *molecular level*) and its effects on responding extended over time (i.e., at the *overall, global,* or *molar level*). The ability of MPR to account for behavior at both levels is, for Mazur, one of its virtues. Another is the ability to account for behavior when it is first acquired as well as its occurrence following acquisition (i.e., in the steady state). Taken together, these virtues of MPR—predictive precision, integration of diverse phenomena, and extensibility to molecular and molar analyses and also to acquisition and steady state—may be considered the gold standard for theories of operant conditioning.

In the remainder of the chapter, we turn to several topics that represent the directions in which the study of operant conditioning has trended in recent years. In almost all cases, the topics are (a) instantiated in research that involves reinforcement schedule methodology and (b) expressible as mathematical models that can be anchored in one way or another to Skinner's founding theory of the three-term contingency.

A SAMPLER OF TRENDS

What Is the Role of Reinforcers?

We previously made reference to the ubiquity of conditioned reinforcers in daily life and to the experimental study of conditioned reinforcement using concurrent chains schedules. A simple question about conditioned reinforcers is whether they have effects similar to those of the primary reinforcers they are associated with—a similarity implied by the joint use of the term *reinforcer*. For example, do the advertisements (one form of conditioned reinforcer) that appear in popular media affect our preferences in the same way as the actual products they portray affect them (i.e., by strengthening or weakening preference)? Or, alternatively, do conditioned reinforcers merely serve as placeholders for the primary reinforcers whose availability they signal? In other words, do reinforcers have a strengthening or a discriminative function?

Davison and Baum (2006) had previously studied "preference pulses" in pigeons using concurrent schedules of reinforcement in which they recorded every peck and every reinforcer. A preference pulse was identified by the persistence of responding to the same schedule following reinforcer delivery (i.e., preferring to remain under that schedule rather than switching to the other schedule that was available). A preference pulse grew larger as the number of responses under the same schedule increased before the subject switched to the other schedule. Davison and Baum used preference pulses to answer a simple question: do conditioned reinforcers produce preference pulses like those that food produces? In their experiment, the conditioned reinforcer was the presentation of a light that illuminated the location where food was delivered to the pigeons. Sometimes the light and the food occurred together. At other times, the light was presented by itself. Preference pulses reliably followed the light alone but were not as large as those produced by food-plus-light presentations, a result consistent with the view that conditioned reinforcers produce reinforcing effects, though not as strongly as primary reinforcers do.

In subsequent conditions, Davison and Baum (2006) pitted food against conditioned reinforcers by altering the degree to which the food ratio between the two schedules was the same as the ratio of conditioned reinforcers. In essence, this procedure allowed them to ask further whether the function of conditioned reinforcers is the same as that of primary reinforcers (i.e., they strengthen the responses that produce them) or whether their function instead is to signal the availability of primary reinforcers (i.e., they have a discriminative function). The results supported the latter view, thus calling into question the longstanding assumption that conditioned reinforcers function as reinforcers per se. Referring to the earlier example, advertisements may not affect our preferences for products so much as provide us with signals of the relative availability of those products: Their function may be that of a discriminative stimulus rather than a consequence. Davison and Baum ended their article with a potentially more provocative question: Do primary reinforcers also exert discriminative effects rather than reinforcing effects? The relevance of this question for the traditional view of the three-term contingency should be clear.

Behavioral Momentum and Conditioned Reinforcers

The preference pulses identified by Davison and Baum (2006) are related to the persistence of behavior following a reinforcer. A similar interest in the persistence of behavior (i.e., resistance to behavioral change) is found in the theory of *behavioral momentum* (Nevin, 1992; Nevin & Grace, 2000). According to the theory, reinforcers affect both response rate and resistance to change. Their effect on response rate reflects a response-reinforcer relation (the familiar strengthening relation). In contrast, their effect on resistance to extinction reflects a stimulus-reinforcer relation (i.e., a Pavlovian relation). Specifically, resistance is affected by the total rate of reinforcement that occurs in the presence of a stimulus. By analogy to physical mass, behavioral mass determines the extent to which changes in response rate (velocity) are affected by changes in reinforcement rate (force). The logic of experimental demonstrations of behavioral momentum depends on the introduction of "disruptors," such as extinction, following the acquisition of stable performance under a standard reinforcement schedule.

A recent experiment by Shahan and Podlesnik (2005) asked whether behavioral momentum also applies to conditioned reinforcers. Pigeons emitted observing responses that produced the stimuli (conditioned reinforcers) associated with the various reinforcement schedules that could be in effect at any one time. Shahan and Podlesnik arranged alternating conditions in which observing responses produced a high rate of conditioned reinforcers or a low rate. They found that observing response rate was higher in the former condition (rich) than in the latter (lean). However, when extinction was used as a disruptor, resistance to change was essentially identical for both conditions. Among the possible explanations the authors offered was that "higher rates of stimulus [conditioned reinforcer] production may not have generated greater resistance to change because the stimuli have their effects on response rates through some mechanism other than strengthening by reinforcement" (pp. 15–16). This possibility aligns nicely with Davison and Baum's (2006) conclusion.

Further Questions About the Three-Term Contingency: The Misbehavior of Organisms and Superstition

Davison and Baum's (2006) and Shahan and Podlesnik's (2005) work raises questions about the adequacy of Skinner's three-term contingency for the analysis of operant behavior. In a probing examination of its adequacy, Timberlake (2004) reflected on Skinner's consummate skill as a shaper of new behavior in animals and the related work of two of Skinner's students, Keller and Marian Breland (later Breland Bailey). With reference to Skinner's successive (and successful) innovations for measuring rats' behavior as outlined in *The Behavior of Organisms*, as well as subsequent clever demonstrations of shaping (rats lowering ball bearings down chimneys, pigeons bowling or playing ping pong), Timberlake pointed up Skinner's ready eye for the present and potential behaviors in an animal's repertoire—what Timberlake styled the "proto-elements" of contingency-shaped behavior. Equally impressive was Skinner's eye for the physical modifications of apparatus that were necessary to produce the desired behavior—"tuning the apparatus." His finesse was underscored by the Brelands' subsequent discovery of instances of misbehavior that emerged unbidden when reinforcement contingencies were applied—behavior gone awry despite the apparently orthodox implementation of shaping techniques. For example, one project involved training a pig to deposit wooden coins in a piggy bank (K. Breland & M. Breland, 1961). When the trainers increased the response requirement, the pigs paused midroute and began to root the coins instead of carrying them to the bank. Such misbehavior persisted despite the loss of the reinforcers that awaited successful deposit.

For Timberlake (2004), successful implementations of the three-term contingency may mask the judicious conjunction of the elements that are operative, including the selection of the species whose behavior will be studied, the selection of the target behavior, the selection of the reinforcer and discriminative stimuli, and the selection of the contingency itself. Because each of these factors is related to each of the others in possibly complex ways, simply aligning arbitrary discriminative stimuli with arbitrary responses and reinforcers in no way assures the occurrence of operant conditioning. Each aspect of the three-term contingency and the relations between them are fitting candidates for further research.

Much the same point was made by Williams (2001), who revisited Skinner's (1948) classic demonstration of so-called superstitious behavior in pigeons. Food was presented to pigeons at regular intervals but was not contingent on behavior. Regardless of the absence of a contingency, Skinner reported that each pigeon soon exhibited its own ritualized behavior and concluded that these emergent behaviors testified to the selective action of the reinforcer. Williams also referred to the classic extension of Skinner's experiment by Staddon and Simmelhag (1971), who found that the behaviors Skinner observed were actually components of a natural repertoire elicited by food. What appeared to have been operant conditioning on the surface turned out to be a case of Pavlovian stimulus-reinforcer relations.

Williams (2001) concluded, as did Timberlake (2004), that, just as each of the separable relations between contingent stimuli and responses in Pavlovian conditioning makes its own contribution to conditioning, so the separable relations between the elements of the three-term contingency must be evaluated for their differential contributions to operant conditioning.

Behavioral Economics: Temporal Discounting and Self-Control

Even though the requisite responses occur under a schedule of reinforcement, delivery of the contingent reinforcers may be delayed (see Catania's, 2005b, reference to the delay-of-reinforcement gradient). What a person prefers right now—say, ice cream—may no longer be preferred if the person has to wait a week for the ice cream or a day or even an hour. Many people, especially children, know the frustration of ordering something they want from an online catalog, only to have to wait for it to arrive in the mail. Being able to obtain it faster somewhere else, even if it is more expensive there, may be preferred to having to wait for it.

The study of choice, including choice between immediately available and delayed reinforcers, eventually prompted an intersection between operant conditioning research and experimental economics known as *behavioral economics*. The shared concepts and terminology include cost-benefit analysis, maximization of utility, marginal utility, budget, and discounting (present vs. future value), among others. Of particular interest has been the shape of the standard discounting function. Traditionally, economics has viewed the curve as falling off rather quickly then leveling off (i.e., discounting tends to be steeper with shorter waits). This phenomenon is referred to as *exponential discounting* in reference to the mathematical form of the discount function. However, operant research with humans and animals has shown that a different kind of function—the *hyperbolic function*—better describes the results of experiments in which individuals make repeated choices between options available now and those available later. The hyperbolic function typically falls more steeply than its exponential counterpart. In addition, hyperbolic functions more readily lend themselves to the reality of preference reversals, as occur, for example, when an individual sets the alarm clock earlier than usual, fully planning to take advantage of additional time that earlier rising will afford. The next morning, however, the folly of the previous night's plan is all too obvious when the alarm clock buzzes; a few more minutes of sleep has a newfound deliciousness, it seems.

Green and Myerson (2004) summarized the experimental support for hyperbolic discounting in the case of choices involving delayed reinforcers as well as that in which the choice is between reinforcers with different odds or probabilities. For example, the discounting functions were steeper for children and young adults than for older adults. Among older adults, they were steeper for those with lower incomes than for those with higher incomes, where education level was held constant. Steeper discounting functions appeared for persons addicted to heroin and alcohol, as well as for smokers than nonsmokers. It is also the case that steeper discounting occurred in pathological gamblers than in only occasional gamblers or nongamblers. Although affirming the primacy of the hyperboloid function for describing temporal discounting, Green and Myerson conceded the real possibility that the processes of discounting for delayed as opposed to probabilistic outcomes may be fundamentally different—two processes rather than a single process—and pointed to the need for further research to resolve the matter.

Rachlin (2000) applied the concept of hyperbolic discounting to delayed outcomes and probabilistic outcomes using a theoretical framework he termed *teleological behaviorism*. On this view, the proper level of analysis is not individual responses or rates of response but patterns of responses extended in time. Moreover, reinforcer delay (measured as units of time) and reinforcer probability (measured as odds) have comparable effects on choice—reinforcers that come sooner are preferred to those that come later, just as those that more likely are preferred to those less likely. These preferences may obtain even when the later arriving or less likely reinforcers are more sizeable than their sooner arriving or more likely counterparts. By behaving in ways that prevent access to the sooner available or more likely alternative—that is, by engaging in what Rachlin called *commitment*—an individual foregoes the reinforcers that would otherwise prevail in favor of those that are richer. This achievement is self-control. It occurs in isolated instances (getting up an hour earlier than usual even on an especially cold morning), or it may become a pattern of behavior that occurs over longer stretches of time, eventually characterizing an individual lifestyle—for example, prudential life as opposed to profligate. Rachlin (2002) also has shown how teleological behaviorism can resolve the inherent contradictions of altruistic behavior.

Behavioral Ecology in the Operant Laboratory

Although methods of operant conditioning may be applied in naturalistic settings (see Baum, 1974a, for ideas on studying pigeons' behavior in one's broken-windowed attic), their typical provenance is a laboratory. However, even there it is possible to examine behavioral phenomena that have been identified previously by ecologists in the wild, including the behavior of groups of subjects. Baum and Kraft (1998) approximated natural patches of food by feeding a 30-member flock of pigeons in ways that constrained both the size of areas in which food was available and the delivery of food to the areas. They varied the travel distance between areas and also the overall rate of food delivery. In a lengthy experiment, they compared the individual allocation of behavior between areas by members of the flock to the aggregated allocation of the flock as a whole.

The analysis of operant behavioral allocation (i.e., of choice between alternative sources of reinforcement) had led to the formulation of the generalized matching law (Baum, 1974b; Davison & McCarthy, 1987). It states that, when two alternatives are available, the proportion of responses directed to one alternative equals the proportion of reinforcers provided by that alternative, subject to a pair of factors known as *bias* and *sensitivity*. Bias refers to the fact that the experimental subject may prefer one reinforcer consistently to another regardless of differences in frequency, amount, or other aspects. Sensitivity refers to the extent to which the subject discriminates the differences between reinforcers. Baum and Kraft (1998) asserted the mathematical correspondence between the generalized matching law and the ideal free distribution, a standard model in behavioral ecology for describing group foraging. In essence, it describes the equilibrium that results when individual foragers switch between food patches in order to optimize individual net gain.

Across the several conditions of their experiment, Baum and Kraft (1998) found that the members of the flock participated to different but consistent extents, perhaps reflecting individual differences in competitive ability. However, there was little if any consistency between individual subjects' responses to shifts in food distribution between the areas, travel time between areas, or the total amount of food available. Instead, what emerged was a collective adherence to the ideal free distribution; it described well the behavior of the flock as an entity.

In a subsequent study, Kraft and Baum (2001) applied the ideal free distribution to the behavior of human subjects in a study analogous to that with pigeons. Instead of moving between food patches, each subject in a 13-person group displayed a colored card (blue or red) to which different numbers of points were assigned and received points based on the number of subjects who displayed the same color. For example, in the 80 to 40 condition, those displaying blue cards divided 80 points and those displaying red cards 40 points. In the 20 to 100 condition, those displaying blue cards divided 20 points and those displaying red cards 100 points. Across the successive trials in each condition, the distribution of cards displayed within the group tended to match the points ratio, thus achieving an equilibrium in which gains were approximately equal for each subject and in accordance with the ideal free distribution.

Animal Timing and Cognition

The experimental efforts to demonstrate similitude between the operant behavior of animals foraging in natural habitats and that of humans playing a card game in the laboratory reflect an interest in complex behaviors shared by humans and nonhumans alike, which include covert or private behaviors—behaviors observable only by the individuals to whom they belong. Timing behavior is one of them. Of course, humans have access to a variety of timekeeping devices but are still able to estimate time in their absence. The experimental analysis of animal timing was part of Skinner's oeuvre. A recent review (Lejeune, Richelle, & Wearden, 2006) of operant methodologies and models of animal timing recognized Skinner's role and his scrupulous avoidance of any reference to internal clocks or other mentalistic time-telling. Of particular moment were Skinner's invention of the fixed-interval (FI) schedule of reinforcement and schedules for the differential reinforcement of response rate, which allowed Skinner and others to study temporal discriminations and additional phenomena related to temporal control (i.e., the control of behavior by intervals of time). More recent behavior-analytic accounts of animal timing have emphasized the role of behavior in mediating temporal discriminations. Lejeune et al. paid close attention to three models in particular: Killeen and Fetterman's (1988) Behavioral Theory of Timing (BeT), Machado's (1997) Learning to Time Model (LeT), and Dragoi, Staddon, Palmer, and Buhusi's (2003) Adaptive Timer Model (ATM). The latter model distinguishes between reinforced behavior and all other behavior and assumes that the temporal control of behavior is reflected in the patterns of alternation between the two types of behavior over time. Alternations occur less frequently as rate of reinforcement declines—a process represented by a decay parameter in the ATM. Under FI schedules, for example, the relatively long pause following a reinforcer is attributed to sustained patterns of other behaviors that give way to the resumption of reinforced behavior in close advance of the next reinforcer.

A similar insistence on accounting for complex behavior in terms of behaviorally defined mechanisms characterizes the study of animal cognition (i.e., animals' private or covert behaviors) within the operant conditioning literature: Any mechanism that is invoked should be readily susceptible to experimental demonstration. Work summarized by Zentall (2001) is illustrative of this process for assuring the parsimony of behavioral theories. Zentall and his students have used the procedure of delayed matching-to-sample to study pigeons' "representation" (a cognitive term of reference) of visual stimuli. In the basic procedure, an arbitrary stimulus (the sample) is presented for a fixed duration. Later, two comparison stimuli are presented, one of which is the original stimulus. The pigeon's task is to select (match) that stimulus from the pair. By manipulating the delay between presentation of the sample and the comparison stimuli, it is possible to produce a retention function, or a function that relates the probability of a correct match to the length of the delay.

Zentall (2001) further explained that the shape of the retention function can be useful in deciding the question of whether the subject's representation is prospective or retrospective: Is the pigeon's behavior during the delay controlled by the sample (retrospective representation) or by the correct member of the comparison pair (prospective representation)? One way to answer the question is to manipulate the type of stimulus that is used—different hues as opposed to different line orientations—for the sample and comparisons in order to determine whether the retention functions differ in corresponding ways. When Zentall and his colleagues arranged such tests, they found conclusive evidence that the representation was retrospective. Varying the sample stimuli affected the retention function, but varying the comparison stimuli did not. Based on such findings, Zentall concluded,

> The advantage of cognitive accounts is that they encourage us to examine behavioral phenomena in a new light and by so doing to conduct novel experiments that often lead to a better understanding of the behavior of organisms, whatever results are obtained. (p. 75)

Behavioral Pharmacology: Drugs as Discriminative Stimuli But Not Necessarily Reinforcers

Branch (2006) made a similar claim in a recent review of the contribution of behavioral pharmacology to the experimental analysis of behavior; behavioral pharmacology is the study of the effects of drugs on behavior. Branch suggested that drugs may serve as discriminative stimuli in the three-term contingency and, in that role, help elucidate the stimulus control of private or covert events. To make the point, Branch first recognized that people typically learn to describe private or covert events from the same people who teach them to describe public events—events mutually available to the individual and to others. For example, a parent who accompanies her or his child to the health clinic for the child's vaccination may express concern at the child's signs of distress after the injection is administered—"I can tell that the needle hurt you." By generalizing from the private events that accompany the parent's verbal behavior on such occasions, the child may discriminate future instances where the expression "That hurts!" is reinforced as a report of covert behavior. Reinforcement is especially likely to occur when the child's self-report is accompanied reliably by overt indicators of pain otherwise.

Because animals do not report private events in the same verbal terms as humans do, experimenters must find alternative means for such reports to occur. Branch (2006) argued that behavioral pharamacologists do so when they administer a drug or a nondrug (such as saline solution) to subjects. Analysis of a subject's overt performances in response to these two types of discriminative stimuli may show differences. All other things being equal, the overt differences can be taken as evidence of the differences in private events that accompany the administration of drug and nondrug, just as a child's wincing and crying following a needle prick aver the self-report, "It hurts!" In fact, Branch proposed the use of procedures in which the presentation of public stimuli that accompany the administration of drug or nondrug is varied systematically. Doing so would allow the experimenter to estimate the relative robustness of private events versus public events.

In addition to serving as discriminative stimuli, Branch pointed out that drugs also reinforce behavior and cited a study by Donny et al. (2003), who trained rats to press a lever to deliver intravenous injection of nicotine. Using a standard procedure, they first trained rats to press one of two levers for food. Eventually lever presses produced food and nicotine. Presses on the second lever (the control lever) had no prearranged outcome. Food delivery was discontinued later. Rats continued to press the lever to deliver nicotine, albeit at a much lower rate than when food was available but still higher than pressing the control lever. When saline subsequently was substituted for nicotine, the rate of lever pressing decreased to become essentially the same as that on the control lever. At this point the results would seem to support the conclusion that nicotine reinforced lever pressing.

However, Donny et al. (2003) were not finished. As an important further control on whether nicotine is a reinforcer, they utilized a second group of rats, each of which was yoked to a subject in the first group—when a rat in the first group received an injection of nicotine, the yoked rat in the second group received an injection of nicotine simultaneously. That injection was not contingent on lever pressing, as it was for the rat in the first group. Nevertheless, the rate of lever pressing was nearly identical by rats in both groups. More importantly, lever pressing declined similarly in both groups when saline was substituted for nicotine and returned to near-original levels when nicotine was reintroduced. These surprising results call into question whether nicotine (and, by extension, other drugs) serves as a reinforcer or rather affects behavior in a different role. For example, nicotine may enhance some other aspect of the situation that is the real reinforcer therein. In any event, Branch (2006) concluded that research in behavioral pharmacology may illuminate issues for the study of behavior in general and not just the effects of drugs on behavior.

Behavioral Neurobiology: A Case Study in Attention-Deficit/Hyperactivity Disorder (ADHD)

The study of drug effects on behavior obviously opens the door to the study of those effects as correlated with co-occurring physiological events in the same subject. This possibility raises the question of the extent to which physiology should play a role in the experimental analysis of behavior. Recently, Timberlake, Schaal, and Steinmetz (2005) summarized Skinner's view of the issue as follows: Neurobiology will eventually provide a unified account of behavior. That account will owe a considerable debt to the experimental analysis of behavior, which will be an invaluable guide to physiologists as they decipher nervous system events and their interrelations: Behavior analysis is propaedeutic to neurobiology, providing grist for the mill, as it were. Or to put it more tendentiously, neurobiology will proceed best by seeking the physiological underpinnings of the phenomena already identified by behavior analysts.

Skinner was not ready to turn behavior analysis over to neurophysiology, as he considered the latter insufficiently mature to tackle the complexities the former was revealing. Timberlake et al. (2006) expressed their impatience with Skinner's insistence on waiting for a readily integrative neurobiology and argued instead that, where bridges between specific behavioral and neurobiological phenomena suggest themselves, they should be pursued. As a case in point, the expansive special issue of *Journal of the Experimental Analysis of Behavior* they edited was meant as a collection of prototypical bridges by which to link the two disciplines.

A theoretical article by Sagvolden, Johansen, Aase, and Russell (2005) provides a further illustration of the prospect. Sagvolden et al. first categorized the clinical symptoms of attention-deficit/hyperactivity disorder (ADHD), all of which are behavioral, with emphasis on the main symptoms of inattentiveness, overactivity, and impulsiveness. They then proceeded to articulate a developmental theory of ADHD that focuses (perhaps ironically) on hypoactivity specifically, lowered activity in three different branches of a brain system featuring the neuromodulator dopamine. Deficits in dopaminergic function alter the function of specific neurotransmitters in brain circuits the authors designate the prefrontal loop, the limbic loop, and the motor loop.

For example, in the limbic loop, brief increases in dopamine activity have been recorded following the presentation of primary and conditioned reinforcers. Eventually such increases are evoked by the stimuli that signal the response-reinforcer contingency (i.e., by discriminative stimuli). Conversely, brief decreases in dopamine activity have been observed when reinforcers are no longer presented following responses during extinction. In addition to explaining how such effects may be mediated at the cellular level, Sagvolden et al. (2005) also described the predictable behavioral effects of the hypofunctioning dopaminergic system. A weakened limbic response to reinforcers may be linked to the symptom of delay aversion or impulsiveness. In other words, the weakened limbic response results in steeper delay-of-reinforcement gradients for individuals diagnosed with ADHD. In such cases, Sagvolden et al. recommended that more robust reinforcers be required as part of therapy for such individuals. They also discussed the implications of their theory for the role that parents, social agencies, and medication play in remitting the symptoms of ADHD.

In an extended commentary on the Sagvolden et al. (2005) article, Catania (2005a) further developed the implications of altered delay-of-reinforcement gradients for the symptoms of ADHD. He argued that steeper gradients reflect the differential selection of rapid response rates (hyperactivity) and the weakened ability of conditioned reinforcers to maintain observing responses (attention deficit), as well as greater impulsiveness. An additional phenomenon may be at work, namely, extinction deficit, which he defined as the failure to observe stimuli that accompany extinction, thereby slowing its achievement. Catania recommended interventions that begin with deliberately short delays of reinforcement and the use of conditioned reinforcers in order to progressively extend the sequence of desirable behaviors that will persist when delays grow longer. He particularly advocated the use of computer games for this purpose.

Donahoe and Palmer (1994; see Donahoe, 2002, for an updated summary) described an alternative approach to the analysis of neurodevelopmentally mediated behavior, such as ADHD. Theirs is a Darwinian approach that draws heavily from the discipline of artificial intelligence and specifically the concept of neural networks. According to this view, the nervous systems of animals, including humans, may be characterized in terms of operational units that selectively process incoming signals and are, in turn, affected by the relative success of the behavior they effect. Donahoe and Burgos (2005) provided a commentary on the Sagvolden et al. (2005) article and on Catania's (2005a) commentary in which they repeated the selectionist theme and applied a selection network simulation in analyzing the differential effects produced by varying the length of reinforcer delay. They showed that, with increasing delays, the decrease in the primary reinforcer's effectiveness could be mitigated by the presentation of conditioned reinforcers during the delay interval—a process reminiscent of that recommended by Catania.

Signal Detection and Reinforcement History

Consideration of the symptoms of ADHD in the context of behavior analysis raises questions about the efficacy of stimuli—discriminative or reinforcing—in the control of behavior. With reference to Skinner's previously cited aversion to cognitive theories, Wixted and Gaitan (2002) offered a middle ground. In their view, cognitive theories have generated research that is to be respected for revealing otherwise unsuspected human abilities. At the same time, cognitive theories may obscure the reinforcement contingencies of which behavior is a function by taking the place of individual reinforcement histories (i.e., by functioning as surrogates for those histories). The authors offer cognitive models, specifically signal-detection models of human recognition memory, as a case in point.

Signal-detection models include parameters for individual sensitivity and individual bias and thus have something in common with the generalized matching law referred to earlier. The models have wide applicability to decision tasks requiring the recognition of a specific event against a background of noise—for example, recognizing a specific face amid a sea of faces. In a typical recognition memory task, subjects first study a list of words then later decide which words on a test list were on the first list. Half of the items on the test list were included in the first list (targets); the other half were not (lures). Four outcomes are possible: The subject correctly recognizes a target (a "hit" in the parlance of signal detection theory); the subject fails to recognize a target (an "omission"); the subject identifies a lure as a target (a "false alarm"); or the subject correctly identifies a lure (a "correct rejection").

Wixted and Gaitan (2002) considered a pair of models for recognition memory. According to one model (familiarity), decisions are based on the subject's relative familiarity with the test items. The other model (likelihood ratio) stipulates that decisions occur on the basis of the odds ratio—the likelihood that an item was drawn from the targets as opposed to the lures. Consequently, it is the more complex of the two models. In further discussion, the authors showed that, as hit rate increases, the familiarity model predicts that the false alarm rate will remain constant. By contrast, the likelihood ratio model predicts that the false alarm rate will decrease and vice versa (the mirror effect). The latter prediction is strongly supported by the recognition memory literature, even though it requires considerably more sophistication on the part of the subject than the familiarity model does. Wixted and Gaitan reviewed additional findings in which the likelihood ratio made unique predictions that experimental results confirmed.

Then their analysis took a provocative turn. They asked whether the complex abilities of humans displayed in recognition memory tasks are biological endowments (i.e., part of a cognitive apparatus that defines humans by nature) or instead might be considered to result from a substantial, individual history of reinforcement for decision making. In an attempt to answer the question, they reviewed data from pigeon experiments involving recognition memory tasks akin to those used with human subjects. The procedure involved two kinds of trials—sample and no-sample. In the former, a sample stimulus appears, followed by a retention interval, then a choice between two test stimuli. A response to one of them produces the reinforcer. In no-sample trials, the trial begins with the retention interval and ends with a response to one of the two test stimuli. Here it is the other stimulus that is correct. In this way, the pigeon discriminates between sample and no-sample trials. Pigeons learn to perform the task with high accuracy and also produce the mirror effect, consistent with the likelihood ratio model. Wixted and Gaitan (2002) reviewed the results from experiments in which there was systematic manipulation of the retention interval across sessions or within sessions. In both cases, they observed a similar pattern of initial disruption of performance followed by a return to criterial performance predicted by the likelihood ratio model—a pattern they directly attributed to the subject's history of reinforcement during the experiment. To drive the point home, the authors demonstrated the mathematical equivalence of the generalized matching law referred to in an earlier section and the likelihood ratio model.

Wixted and Gaitan's (2002) conclusion is imposing: It is unnecessary to account for the results of the pigeon research in terms of cognitive mechanisms; reinforcement history suffices. By extension, similar accounts conceivably could be made for human recognition memory, were sufficient details of individual reinforcement history available. In the authors' words,

> Skinner (1977) was probably right when he asserted that "the mental apparatus studied by cognitive psychology is simply a rather crude version of contingencies of reinforcement and their effects" (p. 9). If so, much of what is interesting about human memory will be illuminated by studying animals whose reinforcement history can be studied in a much more direct way. (p. 305)

SUMMARY

The tension between the experimental analysis of behavior and cognitive psychology may be largely rhetorical (it is a matter of differences in how one talks scientifically about the same data) but still runs deep. Tensions are also discernible within behavior analysis over the status of hypothetical, noncognitive mechanisms of behavior (see, for example, the Book Review and Commentaries section in the July, 2004, issue of *Journal of the Experimental Analysis of Behavior*). Reese (2001) is one of many commentators (see also Critchfield, 2006; Lattal & Chase, 2003; Roediger, 2004) who have argued for a measured approach to these and wider tensions.

The three-term contingency of reinforcement is the touchstone of operant conditioning. DeGrandpre (2000) demonstrated its relevance for postmodern perspectives, including deconstructionism and social constructionism, using drug abuse and its treatment as a case in point. Mattaini (2006) extended the purview of operant conditioning to larger-scale projects for lifting human suffering, including "domestic violence, terrorism, the environmental impact of consumer behavior, substance use, rehabilitation of prisoners, the science of nonviolence, human rights, child labor, racial discrimination, gambling, and HIV/AIDS, to name only a few" (p. 10).

A recent article by Killeen (2004) provides an integrative adieu for our chapter. He introduced stimulus discrimination theory (SDT), with its joint emphasis on sensitivity and bias, and its avatar, signal detection theory. Sensitivity refers to stimulus control of the response and bias to control of the response by the reinforcer; thus, SDT aligns with the three-term contingency of reinforcement. Killeen applied SDT to Skinner's (1957) analysis of verbal behavior, specifically to the categories known as tacts (verbal behavior largely under the control of discriminative stimuli) and mands (verbal behavior largely under the control of reinforcers), and then took up the issue of causal attribution (i.e., judgments about which actions within a context are responsible for a specific consequence). In doing so, he showed that pigeons and humans assign causation similarly and that inferences about the behavioral effects of covert stimuli are premised on the observation of behavior when stimuli are overt. In other words, suppose behavior is observed to change when a new stimulus is added to a previously presented stimulus. A similar behavioral change occurs when drug administration coincides with the previously presented stimulus. Such behavioral correspondence—the same effects whether a new stimulus is presented or a drug is presented—becomes the warrant for inferring that the subjective effect of the drug is akin to adding the other stimulus.

In a tour-de-force, Killeen (2004) also showed that the three-term contingency of reinforcement can be fitted to Aristotle's fourfold of causes—efficient (the discriminative stimulus), material (the response), final (the reinforcer), and formal—the combination of discriminative stimulus, response, and reinforcer in the three-term contingency itself. But Killeen's ultimate interest was to provide a framework for discriminating the causes of behavior on the gamut running from control by reinforcers to control by covert stimuli. What he offered, in the end, was a view of multiple prospective causes linked in a chain. To account scientifically for behavior, the task becomes discriminating the actual causes from the prospective, with reinforcers contingent on doing so. It is a task behavior analysts, of all creatures, can relish.

REFERENCES AND FURTHER READINGS

Baum, W. M. (1974a). Choice in free-ranging wild pigeons. *Science*, *185*, 78–79.

Baum, W. M. (1974b). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, *22*, 231–242.

Baum, W. M., & Kraft, J. R. (1998). Group choice, competition, travel, and the ideal free distribution. *Journal of the Experimental Analysis of Behavior*, *69*, 227–245.

Branch, M. N. (2006). How research in behavioral pharmacology informs behavioral science. *Journal of the Experimental Analysis of Behavior*, *85*, 407–423.

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*, 681–684.

Catania, A. C. (2005a). Attention-deficit/hyperactivity disorder (ADHD): Delay-of-reinforcement gradients and other behavioral mechanisms. *Behavioral and Brain Sciences*, *28*, 419–424.

Catania, A. C. (2005b). The operant reserve: A computer simulation in (accelerated) real time. *Behavioural Processes*, *69*, 257–278.

Critchfield, T. (2006). On the future of behavior analysis: Introduction to the special issue. *Association for Behavior Analysis International Newsletter*, *29*(3), 1–3.

Davison, M., & Baum, W. M. (2006). Do conditioned reinforcers count? *Journal of the Experimental Analysis of Behavior*, *86*, 269–283.

Davison, M., & McCarthy, D. (1987). *The matching law: A research review*. Hillsdale, NJ: Erlbaum.

DeGrandpre, R. J. (2000). A science of meaning: Can behaviorism bring meaning to psychological science? *American Psychologist*, *55*, 721–739.

Donahoe, J. W. (2002). Behavior analysis and neuroscience. *Behavioural Processes*, *57*, 241–259.

Donahoe, J. W., & Burgos, J. E. (2005). Selectionism: Complex outcomes from simple processes. *Behavioral and Brain Sciences*, *28*, 429–430.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Boston: Allyn & Bacon.

Donny, E. C., Chaudhri, N., Caggiula, A. R., Evans-Martin, F. F., Booth, S., Gharib, M. A., et al. (2003). Operant responding for a visual reinforcer in rats is enhanced by noncontingent nicotine: Implications for nicotine self-administration and reinforcement. *Psychopharmacology*, *69*, 68–76.

Dragoi, V., Staddon, J. E. R., Palmer, R. G., & Buhusi, C. V. (2003). Interval timing as an emergent learning property. *Psychological Review*, *110*, 126–144.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Englewood Cliffs, NJ: Prentice Hall.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, *130*, 769–792.

Killeen, P. R. (1988). The reflex reserve. *Journal of the Experimental Analysis of Behavior*, *50*, 319–333.

Killeen, P. R. (1994). Mathematical principles of reinforcement. *Behavioral and Brain Sciences*, *17*, 105–172.

Killeen, P. R. (2004). Minding behavior. *Behavior and Philosophy*, *32*, 125–147.

Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, *95*, 274–295.

Killeen, P. R., & Sitomer, M. T. (2003). MPR. *Behavioural Processes*, *62*, 49–64.

Kraft, J. R., & Baum, W. M. (2001). Group choice: The ideal free distribution of human social behavior. *Journal of the Experimental Analysis of Behavior*, *76*, 21–42.

Lattal, K. A., & Chase, P. N. (Eds.). (2003). *Behavior theory and philosophy*. New York: Kluver Academic/Plenum.

Lejeune, H., Richelle, M., & Wearden, J. H. (2006). About Skinner and time: Behavior-analytic contributions to research on animal timing. *Journal of the Experimental Analysis of Behavior*, *85*, 125–142.

Machado, A. (1997). Learning the temporal dynamics of behavior. *Psychological Review*, *104*, 241–265.

Mattaini, M. (2006). Trends in social issues. *Association for Behavior Analysis International Newsletter*, *29*(3), 10–11.

Mazur, J. E. (2006). Mathematical models and the experimental analysis of behavior. *Journal of the Experimental Analysis of Behavior*, *85*, 275–291.

Nevin, J. A. (1992). An integrative model for the study of behavioral momentum. *Journal of the Experimental Analysis of Behavior*, *57*, 301–316.

Nevin, J. A., & Grace, R. C. (2000). Behavioral momentum and the law of effect. *Behavioral and Brain Sciences*, *23*, 73–130.

Rachlin, H. (2000). *The science of self-control*. Cambridge, MA: Harvard University Press.

Rachlin, H. (2002). Altruism and selfishness. *Behavioral and Brain Sciences*, *25*, 239–296.

Reese, H. W. (2001). Review of *The war between mentalism and behaviorism: On the accessibility of mental processes* by William R. Uttal. *Journal of the Experimental Analysis of Behavior*, *76*, 115–130.

Roediger, R. (2004). What happened to behaviorism? *The APS Observer*, *17*(3), 5, 40–42.

Sagvolden, T., Johansen, E. B., Aase, H., & Russell, V. A. (2005). A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behavioral and Brain Sciences*, *28*, 397–468.

Shahan, T. A., & Podlesnik, C. A. (2005). Rate of conditioned reinforcement affects observing rate but not resistance to change. *Journal of the Experimental Analysis of Behavior*, *84*, 1–17.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1948). Superstition in the pigeon. *Journal of Experimental Psychology*, *38*, 168–172.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1990). Why I am not a cognitive psychologist. *Behaviorism*, *5*, 1–10.

Staddon, J. E. R., & Simmelhag, V. G. (1971). The “superstition” experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, *78*, 3–43.

Timberlake, W. (2004). Is the operant contingency enough for a science of purposive behavior? *Behavior and Philosophy*, *32*, 197–229.

Timberlake, W., Schaal, D., & Steinmetz, J. E. (2005). Relating behavior and neuroscience: Introduction and synopsis. *Journal of the Experimental Analysis of Behavior*, *84*, 305–311.

Williams, B. A. (2001). The critical dimensions of the response-reinforcer contingency. *Behavioural Processes*, *54*, 111–126.

Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning & Behavior*, *30*, 289–305.

Zentall, T. R. (2001). The case for a cognitive approach to animal learning and behavior. *Behavioural Processes*, *54*, 65–78.

38

Social Learning

LAURA TEN EYCK
Children's Medical Center Dallas

Within the discipline of psychology, there are several explanations for the process by which humans and other organisms learn. *Social learning,* or observational learning, is the type of learning that occurs when we acquire new behaviors, or increase the performance of previously learned behaviors, by observing others. At first glance, the topic of social learning may appear so simplistic and logical that it does not require further exploration. After all, who would disagree that we learn many behaviors and skills by observing another person perform them? As we will discover, however, the process of social learning is complex and influenced by a number of factors.

The goal of this chapter is to build a basic understanding of social learning theory. In order to achieve this goal, the chapter begins with a brief review of the development of social learning as a topic of study in psychology, presents an overview of the most widely cited theory of social learning, and explores how researchers use social learning to explain a variety of phenomena in humans.

THE STUDY OF LEARNING

Learning has been a topic of great interest since early in the history of modern psychology. In order to gain a better understanding of the process of social learning, we begin by briefly reviewing two more traditional approaches to the study of learning.

Classical Conditioning

One type of learning studied by early researchers such as Ivan Pavlov, and later John B. Watson, is classical conditioning. Classical conditioning occurs when an organism develops an automatic reflexive response to a neutral event that has no prior meaning. Pavlov demonstrated that he could condition dogs to salivate (a reflex) to the sound of a metronome (a neutral event), even if they did not subsequently receive food. Classical conditioning explains why some of us flinch involuntarily in anticipation of a loud noise, such as a balloon popping, and why our mouths begin to water when we simply think about tart candy or biting down on a lemon slice. Classical conditioning remains a topic of study in the field of learning, and there is a wealth of empirical support for its principles (see Chapter 33, Classical Conditioning, and Chapter 34, Recent Trends in Classical Conditioning). In its most basic form, however, classical conditioning describes the process of learning as it relates to involuntary or reflexive behavior. How then, do we learn voluntary behaviors?

Operant Conditioning

Another type of learning studied by early researchers such as Edward Thorndike and B. F. Skinner is operant or instrumental conditioning (see Chapter 36, Operant Conditioning, and Chapter 37, Recent Trends in Operant Conditioning). Skinner (1953), among others, demonstrated that an organism could acquire a new behavior using reinforcement, or unlearn an old behavior using punishment. In the case of operant conditioning, researchers believe the behaviors are voluntary because the organism chooses to do them in anticipation of a reward, or chooses not to do them to avoid punishment. In this context both rewards and punishments are *consequences* of behaviors. For example, parents can toilet-train children by providing a reward (e.g., praise, brightly colored stickers) each time they successfully use the potty chair. The reward, or consequence, increases the likelihood that a child will choose to use the potty chair in the future. We also can apply the principles of operant conditioning to decrease or eliminate undesirable behavior. For instance, some parents spank their young children when they dart into the street. Researchers believe that a punishment such as a spanking, which is a consequence of darting into the street, decreases the likelihood that a child will choose to engage in the same behavior in the future. Like classical conditioning, operant conditioning remains a widely studied topic in the field of learning, and a great deal of research supports its principles. Nevertheless, operant conditioning does not fully explain the acquisition of all voluntary behavior.

Early Theories of Social Learning

During the first half of the 20th century, many psychologists believed that they could explain most instances of voluntary learning using the principles of operant conditioning (e.g., Skinner, 1953); however, these principles are very specific. From an operant conditioning perspective, an organism spontaneously performs a behavior, which is then immediately followed by a consequence (i.e., reward or punishment). The consequence then influences the likelihood that the organism will perform the behavior in the future. In other words, the reward or punishment must be delivered *directly* to the organism *immediately* after the performance of the behavior for learning to occur. Although this paradigm accurately explains the acquisition of some behaviors, it does not explain why learning often occurs in the absence of a direct and immediate consequence. In an attempt to explain learning that does not involve direct and immediate consequences, several researchers independently began to search for alternative explanations. The resulting research showed that humans and other organisms can in fact learn new behaviors from others they encounter in their social environments.

Miller and Dollard

Miller and Dollard (1941) contended that humans can learn through imitation and described three forms of imitation. The first form of imitation, *same behavior*, occurs when two people independently engage in the same behavior in response to the same stimulus. Miller and Dollard used the example of two people riding the same bus to the same destination. One person did not necessarily learn to ride the bus to the desired destination by imitating the behavior of the other person. Instead, they may have separately read the bus schedule and selected the same bus and route. The second form of imitation, *copying behavior*, occurs when one person learns a new behavior by copying the behavior of another person. Learning through copying often occurs when an external judge provides the person with positive feedback for successfully copying the behavior of a model and negative feedback for unsuccessfully copying the behavior of a model. It is the final form of imitation, *matched-dependent behavior*, that provides the basis for social learning.

Matched-dependent behavior occurs in situations where there is a more experienced model and a less experienced observer. The experience of the model may be related to age, education, or skill level. Miller and Dollard (1941) posited that learning matched-dependent behavior occurs when a less experienced observer imitates the behavior of a more experienced model in order to obtain a similar consequence. To illustrate, a young boy may imitate the behavior of his older sister because he has observed that she receives a bowl of ice cream for dessert when she carries her dinner dishes to the sink. Because the boy has observed his sister receiving a reward, he is cued to perform the same behavior. In this example, the boy *does not immediately or directly receive a reward*. Instead, he is motivated to perform the behavior because he first observed his sister receiving a direct and immediate reward.

Miller and Dollard (1941) tested their model of social learning in a series of experiments using both children and animals, and concluded that learning through matched-dependent behavior is a special form of operant conditioning. Observers use the behavior of the model as a discriminative stimulus to determine when they also should do the behavior. A *discriminative stimulus* is a specific cue that an organism learns to associate with a specific outcome, and was originally identified as part of the operant conditioning paradigm. A pigeon, for example, may learn that the presence of a green light means that pecking a key will result in a food reward and that a red light means that pecking a key will not result in a food reward. In our example, the sister's behavior becomes the stimulus that elicits the same behavior in her young brother because he has learned the association between the action (i.e., carrying the dishes) and the outcome (i.e., ice cream).

Rotter

Rotter (1954) was interested in the process of learning through observation and imitation, and how it relates to the development of personality. His social learning theory of personality is based on the premise that people acquire knowledge about how to behave through significant interactions with their environments and with other people. Further, Rotter viewed social learning as a dynamic process that is driven by past experience, present circumstances, and individual needs.

According to Rotter (1954), there are two factors that determine whether a person is likely to perform a specific behavior in a certain situation: *expectancy* and *value*. Expectancy is the degree to which a person believes he or she will be reinforced for the performance of a behavior. Using our prior example, Rotter would predict that the young boy would be more likely to carry his dinner dishes to the sink if he believes he will be reinforced with a dessert. Value is the degree to which a person views the reinforcer as important. Rotter would predict that the young boy would be more likely to carry his dinner dishes to the sink if he places a high value on dessert. Value is related to individual needs, and researchers believe it motivates the performance of a behavior.

The ideas of Miller and Dollard (1941) and Rotter (1954) represent only a small portion of their larger bodies of work on social learning. Their ideas did, however, spark subsequent attempts to understand the process of social learning.

SOCIAL LEARNING THEORY

Although Albert Bandura was not the first researcher to recognize the role of observation and imitation in the acquisition of behavior, starting in the early 1960s, he and his colleagues began to challenge prior contentions regarding the nature of social learning (e.g., Bandura, 1962; Bandura, D. Ross, & S. A. Ross, 1961). Recall that many psychologists believed social learning is a form of operant conditioning that occurs when an observer is vicariously reinforced or punished for a behavior (e.g., Miller & Dollard, 1941).

Bandura and his colleagues argued that there are instances in which people learn through observation, even when the model does not receive a reward or punishment as a result of performing the behavior (e.g., Bandura, 1977; Bandura & Walters, 1963). In their classic "Bobo doll" experiment, Bandura and his colleagues (Bandura et al., 1961) individually exposed preschool children to one of two adult models while in a playroom setting. While they were engaged in an age-appropriate activity, approximately half of the children saw an adult model playing quietly with a building set in another part of the playroom. The remaining children saw an adult model who behaved aggressively toward a five-foot inflatable doll (i.e., Bobo) in another part of the playroom. The aggressive adult model hit, kicked, punched, and pounded on the doll. The adult model also struck the doll in the head with a mallet and threw it in the air.

Later, the researchers introduced each child to another playroom setting, where he or she had a number of toys from which to choose (Bandura et al., 1961). The experimenters categorized some of the toys as nonaggressive toys (e.g., a tea set, crayons, dolls, and teddy bears); they categorized the remaining toys as aggressive (e.g., a three-foot version of the Bobo doll, a mallet, two dart guns, and a tether ball suspended from the ceiling). During this play session, a hidden experimenter observed and coded each child's behavior.

Children who were initially exposed to the aggressive adult model were significantly more likely to engage in the previously modeled aggressive acts toward the Bobo doll during a 20-minute free-choice play session than children who were initially exposed to the nonaggressive adult model (Bandura et al., 1961). Interestingly, the acquisition of aggressive behavior occurred even though the adult model never received any reinforcement for his or her behavior. The results of this experiment provided clear evidence that social learning is not simply a form of operant conditioning, and that it can occur in the absence of any performance incentives.

Over the course of the next decade, Bandura and others worked to understand the factors that influence the process of social learning (e.g., Bandura, D. Ross, & S. A. Ross, 1963; Dubanoski & Parton, 1971; Grusec & Mischel, 1966; Rosekrans & Hartup, 1967). This body of work ultimately contributed to the development of a comprehensive theory to explain when, why, and how social learning occurs (Bandura, 1977).

Bandura's Social Learning Theory

According to Bandura's (1977) social learning theory, our behavior is influenced by information we gather from *models* (e.g., other humans) in our environments. In other words, we observe a person or people perform a behavior, and then store relevant information about that behavior in our memories so that we can use it during an appropriate situation some time in the future.

Take a moment to consider how many behaviors a person learns by first observing other people. Young boys are not born knowing how to shave; instead, they begin the process of learning this common behavior by first attempting to imitate their fathers or other male models they observe. Many people learn numerous sports, such as baseball and golf, by first watching other people perform the actions associated with them. People learn to drive automobiles, prepare scrambled eggs, and perform surgery in much the same way. We frequently avoid injuries and hazards, both minor and major, because we have the opportunity to watch other people perform behaviors before we attempt to do them ourselves.

From a social learning theory perspective, humans and other organisms acquire many new behaviors through observation (Bandura, 1977; Bandura & Walters, 1963). The theory also accounts for social learning that occurs when the observer has prior knowledge of the behavior. Whether the observed behavior is new or has been experienced previously, the model can influence it greatly; we review how models can influence the behavior of observers next.

The Influence of Models

Models can have varying effects on the behavior of observers. The effects, in part, depend on the observer's prior experience with the behavior and the specific situation in which the behavior will be performed. In the following section, I define each effect, present a real-world example, and discuss supportive, experimental evidence.

Modeling Effect

The modeling effect describes situations in which observers learn to perform a completely new behavior or response (Bandura & Walters, 1963). The first time a young child enrolls in a ballet class, she will probably learn to position her arms and feet in ways that were previously unfamiliar. It is also very likely that the acquisition of these new behaviors will be accomplished through modeling. The instructor will demonstrate each foot and arm position, and the students will attempt to reproduce those positions.

Support for the modeling effect comes from the results of Bandura's classic "Bobo doll" experiment. Recall that nursery school children were exposed to an aggressive adult model who abused an inflatable doll in unique ways (i.e., sitting on the doll and striking it with a wooden mallet). Later, those same children displayed identical aggressive behaviors toward the doll, thus demonstrating behaviors that were not previously part of their behavioral repertoire. The experimenters reasoned that these children, who had no prior experience with the Bobo doll or the specific aggressive behaviors modeled by the adult, learned those aggressive behaviors through modeling. What about situations in which there is prior experience with a behavior or response?

Disinhibitory Effect

The disinhibitory effect occurs when people perform a previously suppressed behavior after observing a model perform a similar behavior without punishment (Bandura & Walters, 1963). The tendency for normally law-abiding citizens to engage in looting, rioting, or destructive behavior is a primary example of the disinhibitory effect. In the chaotic environment created by a rioting mob, it is unlikely that police officers would be able to arrest every person who smashes a store window or steals a television. As a result, people who normally suppress the urge to engage in punishable offenses such as vandalism and theft engage in those very behaviors because they see others do so without experiencing any negative consequences.

Researchers have observed the disinhibitory effect in a number of empirical investigations. Children in one study, for example, viewed cartoons containing aggressive content, after which researchers compared their play behavior to the play behavior of a control group (Lövaas, 1961). Children who viewed the aggressive cartoons chose to play with an aggressive toy more often than did children in the control group. Bandura and Walters (1963) contended that the results of this study demonstrate the disinhibitory effect because exposure to the aggressive behavior of cartoon characters, which were not punished for their actions, increased the performance of aggressive acts in general.

Inhibitory Effect

There are also situations in which the observation of a model serves to reduce the performance of a behavior. The inhibitory effect occurs when an observer decreases or eliminates a previously learned behavior after observing a model experience negative consequences for engaging in the behavior (Bandura & Walters, 1963). Although many people are prone to driving faster than the posted speed limit, nearly all of them will reduce their speed when they observe a fellow driver receiving a ticket from a police officer. Because observers often decrease their behavior without experiencing the negative consequences directly, researchers also describe this effect as *vicarious punishment*.

The development of phobias in some young children may result from exposure to models who experience negative consequences when engaging in a behavior (e.g., Bandura & Rosenthal, 1966). For instance, a young child may develop an intense fear of dogs because she observed her mother express fear and anxiety in the presence of dogs. Because it would be highly unethical to intentionally induce such a phobia in a young child, researchers have found more acceptable ways to test for the existence of the inhibitory effect. Berger (1962), for example, used college students and a more innocuous modeling situation (Expt. 2). The basic experimental design included an observer (student), a male model, and an apparatus that some observers were led to believe could deliver a shock. In addition, each observer and model was connected to a device that measured galvanic skin response (GSR). GSR measures changes in the electrical properties of the skin associated with fear or other arousing emotions.

Each student completed the experiment individually, was paired with the same model, and observed 13 condition-specific pairings (Berger, 1962, Expt. 2). Following the presentation of a dimming light some students observed the male model receive a hand shock, after which he jerked his hand away from the apparatus. Other students observed the male model receive a hand shock following the presentation of a dimming light, but he did not remove his hand from the apparatus. Another group of students observed a male model jerk his hand away from the apparatus after the presentation of a dimming light, but no shock was involved. The remaining students were paired with a male model who sat by the apparatus, and was instructed not to touch it after the light dimmed because it would deliver a shock.

Students who observed the male model being shocked and jerking his hand away from the apparatus demonstrated a greater increase in GSR than did students in the other three groups (Berger, 1962, Expt. 2). Berger concluded that repeatedly observing the model's response to being shocked was sufficient to elicit a fear response in the students, which could explain the development of some phobias.

Eliciting Effect

Finally, there are instances in which the actions of a model can serve to increase the performance of a previously learned but infrequent behavior. The eliciting effect occurs when the observation of a model's behavior leads an observer to increase the performance of the same behavior. People may, for example, increase the frequency with which they volunteer, donate money, or even exercise after observing a model engage in similar actions. The difference between the eliciting effect and the disinhibitory effect may be found in the type of behavior performed and its associated emotions (Bandura & Walters, 1963). Recall that disinhibited behaviors, such as aggression are typically associated with anxiety or fear of punishment. In other words, most people do not engage in aggressive behavior because the consequences will likely be negative; yet, watching a model engage in aggression without reprisal may lead people to behave more aggressively. In contrast, elicited behaviors such as volunteering or exercising are often considered to be prosocial and unlikely to result in anxiety or punishment.

Rosenbaum and Blake (1955) provided empirical support for the eliciting effect. They hypothesized that observing another student (i.e., model) agree to participate in an experiment would increase the likelihood that the observer would also agree to participate in the experiment. Target students exposed to the model who agreed to volunteer for the experiment were significantly more likely to also volunteer for the experiment than were target students exposed to the model who refused (Rosenbaum & Blake, 1955). It is likely that all of the target students had experience with volunteering in some capacity; however, it is also likely that the majority of these students did not volunteer for experiments regularly. In other words, volunteering was a low probability behavior that increased in the presence of a positive model. Bandura and Walters (1963) therefore contended that the results of this study provide support for the eliciting effect.

Summary

Social learning theory identifies four different effects that a model can have on the behavior of an observer. The *modeling effect* occurs when individuals learn a new behavior by observing the behavior of another. The *disinhibitory effect* occurs when individuals perform a previously suppressed behavior because they have observed a model engaging in the behavior without experiencing any negative consequence. The *inhibitory effect*, in contrast, occurs when individuals decrease the performance of a behavior because they have observed a model experience negative consequences when engaging in the behavior. Finally, the *eliciting effect* occurs when the behavior of a model increases the performance of a low-frequency behavior.

Although each of the preceding effects may occur independently, there are also situations in which they may occur simultaneously and in different combinations. Our behavior as humans is often complex, and changes as a result of interactions with our environments. In some instances, we also adapt our behavior as a result of experience. The process of social learning is also highly complex and dynamic. Consequently, Bandura's (1977) theory also addresses the influence that other factors have on the process of social learning, and the degree to which it is successful. The following section outlines those factors.

Processes Influencing Social Learning

Models play an important role in social learning; however, there are other processes that also have a great deal of influence. These processes involve the cognitive and physical aspects of social learning that influence the degree to which it is successful. The four processes are attention, retention, motor reproduction, and motivation (Bandura, 1977).

Attention

Social learning theory is based on the premise that people cannot successfully learn through observation if they do not pay close attention to the important features of a model's behavior (Bandura, 1977). Imagine how difficult it would be to learn a complicated dance step, or how to parallel park an automobile, if you were not paying attention when another person demonstrated the proper technique. Research by Bandura and others identified several factors that influence the degree to which people are likely to attend to, and recall, the relevant elements of a model's behavior. Children are more likely to pay attention to models who can exert some form of control or power over them such as a teacher or parent (e.g., Grusec & Mischel, 1966). People are also more likely to attend to the behavior of models who are more rather than less attractive (e.g., Bandura & Huston, 1961). Attention may also vary based on the reinforcing nature of the model. Children are more apt, for example, to model the behaviors of adults who are warm and nurturing than those of adults who are cold and nonnurturing (Bandura & Huston, 1961). The perceived value of the behavior, and individual differences in the processing of information (Bandura, 1977), can also influence social learning. Attending to the model's behavior, nonetheless, is only the first step in successful observational learning.

Retention

Once individuals have successfully paid attention to the important elements of a model's behavior, they must have the ability to remember the behavior accurately enough to perform it during some future situation (Bandura, 1977). In other words, paying close attention to the details of a model's behavior is futile unless it can be accurately remembered and retrieved from memory. Bandura argued that when people gather information about the behavior of another person through observation, they do not necessarily store an exact replica of the behavior or action in their memories. Instead, individuals store a *symbolic representation* of the information, which they may then use as a guideline for similar situations in the future.

From the perspective of social learning theory, there are two different representational systems used during the course of observational learning (Bandura, 1977). The first representational system is imaginal. When attempting to recall the behavior of a model, Bandura contended that we are often able to bring to mind an image of that behavior. The image of a modeled behavior becomes especially embedded in memory after an individual is exposed to the behavior repeatedly, and is easily recalled when memories of that behavior are activated. If experienced golfers read or hear the phrase "teeing off," a majority of them would likely report that they could immediately imagine themselves standing with their feet at shoulder's width, gripping the club with one hand above the other, lining up the head of the club with the golf ball, pulling the club back over the shoulder, and then swinging through to strike the ball with the club. The imaginal representational system is particularly valuable in the early stages of development, when children have not yet acquired the necessary language skills. It is also quite useful when attempting to perform behaviors that are not easily learned via verbal or written instructions, such as riding a bicycle or knitting a scarf.

The second representational system used in the process of observational learning is verbal. Bandura (1977) contended that many behaviors we learn observationally are stored in our memories as a series of verbal codes. Rather than drawing on the image of oneself hitting a golf ball each time, for example, one could store the procedure as a verbal code: "feet, grip, align, pull back, swing!" Verbally coding the memory of a behavior is comparable to using a form of cognitive "short hand," which allows us to store and retrieve information quickly and easily.

Evidence suggests that the use of symbolic imagery and verbal codes associated with behaviors enhances the process of observational learning for both children and adults (e.g., Bandura, Grusec, & Menlove, 1966; Bandura & Jeffery, 1973). Additionally, mental or actual rehearsal aids in the retention and retrieval of observationally learned behaviors because the opportunity to perform such behaviors is not always available or appropriate. To illustrate, people may learn the Heimlich maneuver by watching a model perform it during a film or video; however, it would most certainly be inappropriate for people to practice the technique, and therefore strengthen their memories for the task, by grabbing an unsuspecting victim around the midsection while thrusting clenched fists into his or her diaphragm. Instead, people could either rehearse the actions mentally, or they could enroll in a class and practice with their classmates. In either case, they would increase the probability of accurately recalling the proper technique during an emergency.

Motor Reproduction

Once an individual has attended to the important aspects of a model's behavior, and developed an accurate, retrievable symbolic representation of the behavior, he or she may then attempt to perform the behavior. In order to perform an observationally learned behavior successfully, the individual must possess the necessary motor skills and physical capabilities (Bandura, 1977). Some motor skills and physical capabilities are simply the result of development and maturation. For example, young children must first be physically capable of standing, balancing, and coordinating motor movements before they can successfully ride a two-wheeled bicycle. Other motor skills are the result of experience or practice. Elite figure skaters must first master the basics of ice-skating such as balancing, turning, and stopping before they can successfully learn jumps, spins, or complicated tricks. Consequently, it is possible to pay attention to, and accurately recall, a model's behavior and yet fail when attempting to do the behavior because one lacks physical ability.

Motivation

The final process that influences observational learning is motivation. There are two issues related to motivation in the process of social learning. First, people are more likely to reproduce the actions of a model if they perceive that they will enjoy a similar outcome or gain a similar reward (Bandura, 1977). If people do not believe that learning the behavior is important, or they do not find the potential reward appealing, then they will not likely be motivated sufficiently to learn the behavior. Thus, a lack of interest could lead to poor attention or retention, which could translate into an intentional failure to learn the modeled behavior.

Second, there are instances in which social learning has occurred, but the observer was not sufficiently motivated to perform the behavior. A child, for example, could learn how to open a complicated cabinet latch by watching his mother. During these observations, the cabinet might contain items that are of little interest to the child such as cans of pet food. Consequently, the child never attempts to open the cabinet and his mother likely believes that he is unable to do so. What if, however, the mother decides to begin storing the child's favorite candy in the cabinet? He may then decide to open the cabinet latch and help himself to the candy. Bandura (1977) would argue that although the child was capable of opening the cabinet, he was not motivated to do so until it contained an incentive that he desired. The classic work of Tolman and Honzik (1930) and subsequent researchers (e.g., Colwill & Rescorla, 1985), who demonstrated that performance of a newly learned behavior did not occur until the organism was sufficiently motivated to act, supports this contention. Researchers have described this type of learning as *latent learning* because there is a delay between learning and performance.

Summary

Social learning theory identifies four different processes that affect learning through observation (Bandura, 1977). Attending to the relevant details of a model's behavior is a critical first step in the process of social learning. It is also important to store those details in memory and have the ability to recall them accurately at an appropriate time in the future. Next, one must possess the physical ability and motor skills to reproduce the behavior. Finally, individual levels of motivation can affect the process of social learning, both in the acquisition of behaviors and their subsequent performance.

The Role of Cognition in Social Learning Theory

Bandura's (1977) social learning theory and its predecessors (e.g., Miller & Dollard, 1941; Rotter, 1954) share an important feature that sets them apart from the early theories of learning that were based on the principles of operant and classical conditioning. What made these early theories of social learning different is that they considered the role of cognition in the acquisition of new behaviors.

Recall that classical conditioning explains the acquisition of involuntary or reflexive conditioned responses. Pavlov's dogs did not likely make a conscious effort to begin salivating after hearing the metronome, nor did they have the ability to prevent salivation at will. Instead, salivation became a conditioned reflex after repeated pairings of the buzzer and food. In short, researchers believed that there were no thought processes involved in learning classically conditioned responses.

Also recall that operant conditioning explains the acquisition of voluntary responses, and that changes in behavior result from the delivery of rewards and punishments. In its purest form, researchers originally thought that only external factors influenced operant conditioning (see Skinner, 1953); thinking had no role in learning. Researchers such as Skinner argued that our thoughts are unreliable, subjective, and highly inaccurate; consequently, he believed that a comprehensive explanation of learning should focus only on how environmental factors influence behavior. From this perspective, a child who is repeatedly rewarded with money for cleaning his or her room forms a direct association between the money and the action, without any thoughts about the causal relation between the two variables.

When Bandura began developing his theory of social learning, he was especially interested in the role that thoughts play in the process of learning (e.g., Bandura & Walters, 1963). He disagreed with Skinner's (1953) view that thought has no role in the process of learning, instead arguing that our thoughts are a critical component in the regulation of our actions (Bandura, 1977). In the previous example, Bandura would argue that after repeatedly rewarding a child for cleaning his or her room, the child does acquire knowledge about the causal relation between the variables. Specifically, the child is able to reliably predict that a clean room will result in a monetary reward. Bandura contended that we think about our past behavior and its consequences, which in turn instructs our future behavior.

More recently, theorists have referred to Bandura's social learning theory as social cognitive theory, which better reflects the importance of cognitions in the learning process (Bandura, 1986, 1992). Further, researchers have expanded and applied his original work to a diverse range of topics. The final section highlights some of those applications.

Applications of Social Learning Theory

By using a social learning framework, researchers have explained a great number of psychological and social phenomena. Rather than reviewing each of these explanations, the following section includes three diverse examples of how learning through observation can potentially change behavior.

Aggression

Aggression is one of the most widely researched topics in the field of psychology. Many researchers, including Bandura, believe that aggression is a socially learned behavior. Recall that Bandura et al. (1961) demonstrated that young children learned to behave aggressively toward the Bobo doll after watching an adult model punch, kick, and generally abuse the inflatable doll. In a subsequent study, the same researchers demonstrated that young children also learned to behave aggressively toward Bobo after watching a film of an adult modeling aggressive behavior (Bandura et al., 1963). The results of these studies were highly controversial and sparked a debate about the influence of television, films, and other media on children's development and related behavior that continues today. Researchers, for example, expanded the scope of this original research by demonstrating that viewing violence on television and in films increases violent and aggressive behavior in children (e.g., Josephson, 1987; Leyens, Camino, Parke, & Berkowitz, 1975). More recently, there is also evidence that playing violent video games affects aggressive and delinquent behavior in children (e.g., Irwin & Gross, 1995) and college students (e.g., Anderson & Dill, 2000).

Gender Roles

Gender roles are the patterns of behavior considered appropriate for male and female members of a given culture. In other words, they provide descriptions of how boys and girls are supposed to act. The types of occupations that men and women typically pursue are defined by gender roles. For example, a common gender role for women in our society is "secretary" or "nurse," whereas a common gender role for men is "construction worker" or "doctor." As an adult, you may believe that such stereotypes are no longer accurate representations of our culture; but research indicates that children tend to endorse the same gender roles today as they did in the past (e.g., Reid, 1995). Further, these gender-specific occupations often continue to be dominated by the gender with which they are associated. Gender roles also provide information about the characteristics that men and women typically possess. For instance, society typically expects men to be tough and less emotional than women. How do children acquire information about gender roles?

Social learning provides one of the most logical explanations for the acquisition of gender roles. We learn to "act like a man" or "be a lady" from models in our environments, such as our parents, siblings, relatives, and peers (Bandura, 1986; Mischel, 1966). Little boys often mimic their fathers by pretending to shave, pushing a toy lawn mower, or using a plastic hammer. Little girls often mimic their mothers by pretending to apply makeup, pushing a toy baby buggy, or playing dress up. These descriptions may certainly seem stereotypic and outdated, and you can probably come up with a number of examples that dispute them. Nonetheless, there is ample evidence to suggest that young children continue to engage in and endorse many of the same gender roles that were prevalent in the past (e.g., Franken, 1983; Helwig, 1998; Reid, 1995).

Parents or other primary caregivers are some of the first models that young children observe. It is therefore likely that the transmission of gender roles begins early in a child's life. Research indicates that parents play with male and female children in a different way (Smith & Lloyd, 1978), and that by the age of three most children show distinct preferences for "gender-appropriate" toys, which are most likely provided by parents, among others (e.g., girls prefer dolls, boys prefer toy cars; Huston, 1983). Older siblings may also serve as models for gender roles. In one study, researchers examined the relation between gender role qualities in firstborn and second-born siblings during a three-year period (McHale, Updegraff, Helms-Erikson, & Crouter, 2001). Interestingly, second-born children reported gender role attitudes that were similar to those reported by their firstborn siblings, and these attitudes increased in similarity over time.

The Treatment of Phobias

Recall that some researchers believe that the development of some phobias results from social learning. It occurs when an individual observes another person experience negative consequences after an encounter with a specific object, such as a dog or a spider (e.g., Bandura & Rosenthal, 1966; Berger, 1962). Using the principles of social learning, many researchers have demonstrated that modeling can also reduce or eliminate phobias.

In one study, dog-phobic children between the ages of three and five were assigned to one of three treatment conditions (Bandura & Menlove, 1968). Some children watched a film in which a single peer model interacted with a dog in progressively more intimate ways. Other children watched a film in which various models interacted with many different dogs in progressively more intimate ways. The remaining children watched movies that did not feature any animals. After watching their respective films, children interacted with an actual dog in a series of progressively more intimate tasks. The researchers designed the tasks so that each increase in intimacy was related to an increase in their aversiveness. In the least intimate task, for example, children approached a playpen in which the dog was confined and simply looked at the dog. In the most intimate task, children climbed into the playpen with the dog and petted its stomach. Children in both modeling conditions demonstrated significant decreases in their dog-avoidance behavior when compared to children in the control condition. Similar studies have also used modeling to reduce phobias related to snakes (e.g., Ritter, 1968), water activities, spiders (e.g., Denney, Sullivan, & Thiry, 1977), and needle injections (e.g., Trijsburg, Jelicic, van den Broek, & Plekker, 1996).

SUMMARY

Social learning occurs when we acquire new behaviors, or increase the performance of previously learned behaviors, after observing the actions of a model. Although Bandura (1977) started developing his theory of social learning more than 35 years ago, his work continues to influence psychologists and their research. Further, he and his colleagues have expanded the original theory to address some of today's most important issues such as health promotion (Bandura, 2004) and recovery from trauma (Benight & Bandura, 2004).

One of the most important contributions that Bandura made to the field of learning was acknowledging the role of cognition in acquisition and performance of behavior. It seems impossible to understand the process of human learning without taking into consideration the complex nature of our experiences. Our thoughts about an observed behavior largely determine whether we will learn it, or choose to perform the behavior in the future. If we do perform the behavior, our experience with it will then determine whether we choose to do it again.

Researchers have used the principles of social learning theory to explain a variety of psychological and behavioral phenomena, and to improve the quality of human life. Learning through observation pervades nearly every aspect of our existence, and is an integral part of our development and socialization.

REFERENCES AND FURTHER READINGS

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78,* 772–791.

Bandura, A. (1962). Social learning and imitation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (pp. 211–274). Lincoln: University of Nebraska Press.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Upper Saddle River, NJ: Prentice Hall.

Bandura, A. (1992). Social cognitive theory. In V. Ross (Ed.), *Six theories of child development: Revised formulations and current issues* (pp. 1–60). London: Jessica Kingsley Publishers.

Bandura, A. (2004). Health promotion by a social cognitive means. *Health Education and Behavior, 31,* 143–164.

Bandura, A., Blanchard, E. B., & Ritter, B. (1969). Relative efficacy of desensitization and modeling approaches for inducing behavioral, affective, and attitudinal changes. *Journal of Personality and Social Psychology, 13,* 173–199.

Bandura, A., Grusec, J. E., & Menlove, F. L. (1966). Observational learning as a function of symbolization and incentive set. *Child Development, 26,* 499–506.

Bandura, A., & Huston, A. C. (1961). Identification as a process of incidental learning. *Journal of Abnormal and Social Psychology, 63,* 311–318.

Bandura, A., & Jeffrey, R. W. (1973). Role of symbolic coding and rehearsal processes in observational learning. *Journal of Personality and Social Psychology, 26,* 122–130.

Bandura, A., & Menlove, F. L. (1968). Factors determining vicarious extinction of avoidance behavior through symbolic modeling. *Journal of Personality and Social Psychology, 8,* 99–108.

Bandura, A., & Rosenthal, T. L. (1966). Vicarious classical conditioning as a function of arousal level. *Journal of Personality and Social Psychology, 3,* 54–62.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63,* 575–582.

Bandura, A., Ross, D., & Ross, S. A. (1963). Imitation of film-mediated aggressive models. *Journal of Abnormal and Social Psychology, 66,* 3–11.

Bandura, A., & Walters, R. H. (1963). *Social learning and personality development.* New York: Holt, Rinehart & Winston.

Benight, C. C., & Bandura, A. (2004). Social cognitive theory of posttraumatic recovery: The role of perceived self-efficacy. *Behaviour Research and Therapy, 42,* 1129–1148.

Berger, S. M. (1962). Conditioning through vicarious instigation. *Psychological Review, 69,* 450–466.

Colwill, R. M., & Rescorla, R. A. (1985). Postconditioning devaluation of a reinforcer affects instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 11,* 120–132.

Denney, D. R., Sullivan, B. J., & Thiry, M. R. (1977). Participant modeling and self verbalization training in the reduction of spider fears. *Journal of Behavior Therapy and Experimental Psychiatry, 8,* 247–253.

Dubanoski, R. A., & Parton, D. A. (1971). Effect of the presence of a human model on imitative behavior in children. *Developmental Psychology, 4,* 463–468.

Franken, M. W. (1983). Sex role expectations in children's vocational aspirations and perceptions of occupations. *Psychology of Women Quarterly, 8,* 59–68.

Grusec, J., & Mischel, W. (1966). Model's characteristics as determinants of social learning. *Journal of Personality and Social Psychology, 4,* 211–215.

Helwig, A. A. (1998). Gender-role stereotyping: Testing theory with a longitudinal sample. *Sex Roles, 38,* 403–423.

Irwin, A. R., & Gross, A. M. (1995). Cognitive tempo, violent video games, and aggressive behavior in young boys. *Journal of Family Violence, 10,* 337–350.

Josephson, W. L. (1987). Television violence and children's aggression: Testing the priming, social script, and disinhibition predictions. *Journal of Personality and Social Psychology, 53,* 882–890.

Leyens, J. P., Camino, L., Parke, R. D., & Berkowitz, L. (1975). Effects of movie violence on aggression in a field setting as a function of group dominance and cohesion. *Journal of Personality and Social Psychology, 32,* 346–360.

Lövaas, O. I. (1961). Effect of exposure to symbolic aggression on aggressive behavior. *Child Development, 32,* 37–44.

McHale, S. M., Updegraff, K. A., Helms-Erikson, H., & Crouter, A. C. (2001). Sibling influences on gender development in middle childhood and early adolescence: A longitudinal study. *Developmental Psychology, 37,* 115–125.

Miller, N. E., & Dollard, J. (1941). *Social learning and imitation.* New Haven, CT: Yale University Press.

Mischel, W. (1966). A social learning view of sex differences in behavior. In E. Maccoby (Ed.), *The development of sex differences.* Palo Alto, CA: Stanford University Press.

Reid, G. M. (1995). Children's occupational sex-role stereotyping in 1994. *Psychological Reports, 76,* 1155–1165.

Ritter, B. (1968). The group desensitization of children's snake phobias using vicarious and contact desensitization procedures. *Behaviour Research and Therapy, 6,* 1–6.

Rosekrans, M. A., & Hartup, W. W. (1967). Imitative influences of consistent and inconsistent response consequences to a model on aggressive behavior in children. *Journal of Personality and Social Psychology, 7,* 429–434.

Rosenbaum, M., & Blake, R. R. (1955). Volunteering as a function of field structure. *Journal of Abnormal Social Psychology, 50,* 193–196.

Rotter, J. B. (1954). *Social learning and clinical psychology.* Englewood Cliffs, NJ: Prentice Hall.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Smith, C., & Lloyd, B. (1978). Maternal behavior and perceived sex of infant: Revisited. *Child Development, 49,* 1263–1265.

Tolman, E. C., & Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology, 4,* 257–275.

Trijsburg, R. W., Jelicic, M., van den Broek, W. W., & Plekker, A. E. M. (1996). Exposure and participant modeling in a case of injection phobia. *Psychotherapy and Psychosomatics, 65,* 57–61.

39

Stimulus Equivalence

Thomas S. Critchfield and Daniel Fienup
Illinois State University

In stimulus equivalence, untaught abilities arise spontaneously out of a few that are learned through direct experience. Students of psychology often wonder where the most creative and flexible aspects of thought come from; stimulus equivalence may provide a partial explanation. Until recently, scientists had a limited understanding of the experiences that create stimulus equivalence. Contemporary research on the building blocks of stimulus equivalence provides insights into some complex everyday problems, including that of how to teach for maximum benefit. This chapter defines stimulus equivalence, provides some tips for understanding the rather technical methods that are used to study it, and illustrates some applications of the insights that are derived from the research. The chapter also shows how stimulus equivalence is related to some well-known psychological phenomena such as operant learning, concept learning, and knowledge networks in long-term memory.

INTRODUCTION

Among our varied psychological gifts, perhaps most valuable is the ability to "go beyond the information given"—that is, to derive more knowledge than has been learned through direct experience. *Stimulus equivalence* describes such an ability.

Starting about a century ago, John Dewey (1933), whom many consider a father of contemporary educational thought, placed "going beyond the information given" at the center of his views on education. In mastering a new academic subject, he said, we begin by memorizing new facts, but our potential intellectual power goes far beyond that. The goal of education, in fact, is to promote "reflective thinking" that integrates and extends what had been learned through memorization. In a now-famous example of "reflective thinking," Dewey noted that a student who learns that *All men are mortal* and that *Socrates is a man* should be able to derive—without any further instruction—that *Socrates is mortal*.

Dewey's syllogism captures the spirit of stimulus equivalence, a derivation of untaught relations from what has been learned through direct experience. A "relation" can be thought of as a psychological connection between events. Stimulus equivalence exists when we treat relate dissimilar stimuli and treat them as functionally (psychologically) interchangeable. As will be shown later, to scientists "responding similarly" is defined in a very specific way. In general terms, however, think about how "responding similarly" might work in an example that closely parallels Dewey's. Imagine three stimuli: the spoken word "dog" (A), a Doberman pinscher (B), and the printed word *dog* (C). To our senses, these stimuli have little in common, but with the right past experiences we react to them in much the same way. For instance, a glimpse of a Doberman, a printed sign stating *Beware of the dog*, and the spoken warning, "Watch out for that dog!" all can prompt the same precautionary measures.

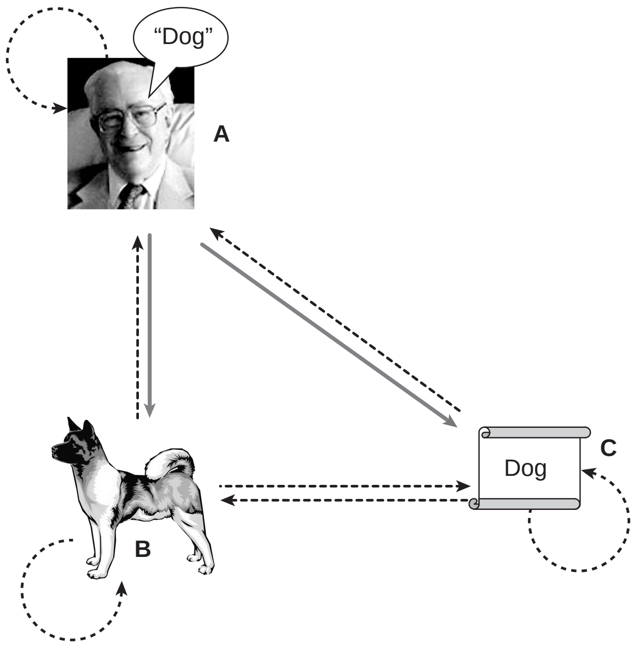

Figure 39.1 A simple stimulus equivalence class incorporating three stimuli. Relations that were learned by direct experience are shown as solid arrows. Those that emerged, untaught, are shown as dashed arrows.

Experience sometimes explicitly teaches us to treat nonidentical stimuli as equivalent—think of the vocabulary exercises involved in learning a second language. "Reflective thinking," however, implies *emergent* equivalence. Assume that a child has been explicitly taught to associate the spoken word "dog" (A) with a photograph of a Doberman (B), a relation that can be expressed as A→B (the arrow notation implies that when A, "dog," is present, B prompts some response, such as pointing to a picture of a Doberman [rather than, say, a wombat]). The child also has learned to associate "dog" with the printed word *dog* (A→C). These explicitly taught abilities are shown as solid lines in Figure 39.1. With these abilities in place, other abilities can emerge without further teaching. Most notably, the child should now relate the picture of the Doberman to the printed word *dog* (A→C), even though these two stimuli have never been experienced together.

An "Engineering" Challenge

As Dewey's syllogism suggests, philosophers and psychologists have long recognized that stimulus equivalence can occur. Many theorists have not been very clear about what experiences are needed to promote the "reflective thinking" that unleashes our full intellectual power. A major contribution of stimulus equivalence research has been to specify some of the conditions under which new abilities reliably emerge—untaught—from those learned through direct experience. Among other things, such knowledge allows "reflective thinking" to be created at will, as the following example illustrates.

Tracy Zinn (2002), a graduate student, used stimulus equivalence to address a learning problem that her professor encountered while teaching a college course on drugs and behavior. Students often became confused because commercial medications have both a brand name and a generic name (see Chapter 17, Drugs and Behavior). For example, Valium® is a brand of the generic antianxiety drug diazepam, and Thorazine® is a brand of the generic antipsychotic drug chlorpromazine. Because many drug names are difficult to pronounce, some students also had trouble connecting spoken drug names with their written equivalents. Overall, students spent so much time trying to learn drug names that they were unable to focus on more interesting features of psychoactive drugs. Zinn showed that instruction based on stimulus equivalence can help people learn drug names with less effort.

Zinn's students worked individually on a computerized instructional module that taught selected relations between trade and generic names of 32 medications in four different drug classes. For each drug, a Stimulus Equivalence Group was taught the relations shown as solid arrows in Figure 39.2 (left side). They learned to recognize written trade and generic names when given the spoken versions of those names, and also to match the written trade name to the spoken generic name. Afterward, the students were tested on three emergent relations among these names (dashed arrows in Figure 39.2).

A "Complete Instruction" Group worked on a different computer module, one that explicitly taught *all* of the relations shown in Figure 39.2 (left panel), before proceeding to the same posttest as the Stimulus Equivalence Group. Not surprisingly, this group took about twice as long to complete their instruction. As shown in the right panel of Figure 39.2, both groups mastered the three relations that were taught explicitly to the Stimulus Equivalence Group. The groups also did equally well on the three relations that the Stimulus Equivalence Group was *not* taught.[1] Thus, this study showed how stimulus equivalence can contribute to instructional efficiency. Both groups succeeded, but for the Stimulus Equivalence Group, "reflective thinking" substituted for a large portion of the instructional effort required to build mastery in the other group.

Formal Definition of Stimulus Equivalence

Untaught abilities like those described above can arise spontaneously once an individual has mastered a few

[1] These outcomes did not occur simply because the material was easy. A third group practiced all of the relations, as per the Complete Instruction Group, but in the same number of trials as completed by the Stimulus Equivalence Group. They scored about two letter grades worse than the other two groups.

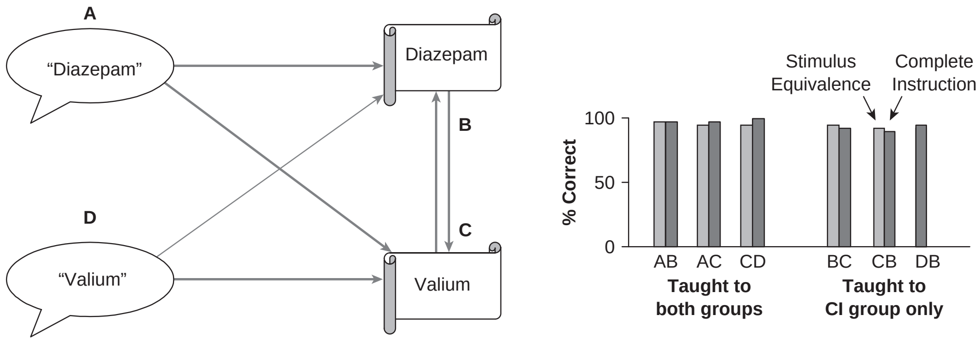

Figure 39.2 How some students efficiently mastered the relations among drug names. *Left side:* Relations acquired through direct experience are shown as solid arrows. Those that emerged, untaught, are shown as dashed arrows. *Right side:* Students taught using stimulus equivalence learned as well as students taught more traditionally, but did so after only about half of the effort. See the text for further details.

SOURCE: Based on Zinn, 2002.

overlapping relations among stimuli. Up to three kinds of untaught abilities (dashed lines in Figure 39.1) can emerge. Scientists label these abilities using terms borrowed from mathematical set theory (for the logic behind this, see Sidman, 1994), and when all three are present, a *stimulus equivalence class* is said to exist.

The emergent ability that best captures the spirit of "reflective thinking" is called *transitivity*. It involves relating stimuli that, although never previously paired in direct experience, share a common associate. In Figure 39.1, direct experience has not taught that the printed word *dog* applies to the Doberman photograph (B→C), or vice versa (C→B), but they may be related by virtue of their common connection to the spoken word "dog" (A).

Psychology students often wonder about the essential ingredients of human nature. Quite possibly, transitivity is one of them. Nonhuman animals show several kinds of emergent abilities, but to date it is not clear whether transitivity is within their reach. Marine Biologist Ronald Schusterman and his colleagues (2003) believe that they have seen transitivity in sea lions, although not everyone agrees with their conclusions. This debate addresses important questions about what it means to be human, but for now it should not distract us from a more fundamental point. As John Dewey lamented, without the right learning experiences, even the brightest among us may not show transitivity spontaneously.

Stimulus equivalence also implies the existence of two other kinds of relations that often are taken for granted in everyday experience. *Symmetry* is the inverse of an explicitly learned relation between stimuli. Having learned that "dog" refers to a Doberman (A→B), an individual may infer that the Doberman is labeled as "dog" (B→A). In *reflexivity* (circular arrows in Figure 39.1), an individual recognizes a stimulus as itself. Reflexivity is required in a common child's game in which a picture (e.g., of a Doberman) must be matched to an identical picture among several pictures (Doberman, wombat, dolphin) spread out on a table. Although not illustrated in the figure, reflexivity also is at the core of recognition memory (see Chapter 46, Memory), which would be required in the child's game if the first Doberman picture were put away before presenting the choices on the table.

At first blush, symmetry and reflexivity seem uninteresting because typical adults so often show these abilities effortlessly, and not all stimulus equivalence research studies focus on them (e.g., note that the study on drug names tested only for transitive and selected symmetrical relations). Yet symmetry and reflexivity are part of the definition of stimulus equivalence for two reasons. One reason is logical: In mathematical set theory, which provides our vocabulary for talking about the relations involved in "reflective thinking" (for more about this topic, see Sidman, 1994), equivalence exists among members of a group of stimuli only if reflexivity and symmetry also exist. The other reason is practical: These abilities do not *always* emerge spontaneously. For example, in the case of symmetry, an English-speaking student of the German language may grasp that *dog* means *hund* before mastering the translation in reverse. In the case of reflexivity, consider a young medical student who, having felt an abnormal mass in cancerous breast tissue for the first time, now struggles to identify an identical mass in a different breast. As these examples suggest, for typical adults, symmetry and reflexivity may be most problematic when unfamiliar stimuli are involved. These relations also can be problematic for very young children, persons with intellectual disabilities, and nonhumans. Such cases create enough uncertainty that many stimulus equivalence studies, even with typical adult participants, test for symmetry and reflexivity, just to be sure they occurred.

Inheritance of Function

The stimuli in an equivalence class may be thought of as psychologically interchangeable. Before coming together in an equivalence class, some stimuli may affect us in very specific ways (e.g., they cause fear) while others are unfamiliar and thus essentially meaningless. When an equivalence class forms, the stimuli may each inherit any psychological functions served by other class members. Two examples to illustrate this *inheritance of function* follow.

Psychologist Dermot Barnes-Holmes and his colleagues have shown how inheritance of function might influence consumer brand-name product preferences (D. Barnes-Holmes, Keane, Y. Barnes-Holmes, & Smeets, 2000). In an early part of the experiment, two equivalence classes formed. One paired nonsense syllables (e.g., VEK, YOF) with words that arouse positive emotions (e.g., “holidays”); the other paired different nonsense syllables with words that arouse negative emotions (e.g., “cancer”). Through inheritance of function, the nonsense syllables, originally psychologically neutral, now would be expected to elicit emotions. In a later part of the experiment, the nonsense syllables were paired with brand names of two carbonated soft drinks. Thus, brand names were related to emotional stimuli, but only indirectly. Finally, participants drank from bottles with different brand-name labels but identical contents. If the flavor of the drinks were all that mattered, there would be no reason to prefer either brand. Instead, the “consumers” clearly preferred the brand that was indirectly associated with pleasant words, an outcome that makes sense in terms of inheritance of emotional functions.

Many kinds of psychological functions are known to spread through equivalence classes (Sidman, 1994), but to illustrate how inheritance of function might be clinically relevant, for our second example let’s stick with the triggering of emotions. Several clinical disorders incorporate learned fears of harmless stimuli. What if one of the stimuli in the equivalence class of Figure 39.1 became a conditioned elicitor of fear? A team led by clinical psychologist Michael Dougher (1998) has found that learned fear can spread through equivalence classes. To get a sense of how their experiments worked, look at Figure 39.3. The three stimuli in the top left box, including the spoken word “dog,” probably do not elicit fear under normal circumstances. Imagine, however, that a child’s mother tells a terrifying story (indicated by a lightning bolt) about the time she was attacked by a neighbor’s dog. The story elicits fear in the child, and, through a possible case of rapid classical conditioning (see Chapter 33, Classical Conditioning), the spoken word “dog” now does as well. Through inheritance of function (middle box), pictures of dogs and the printed word *dog* should now elicit fear, too, even though they were not part of the conditioning episode. This effect—which could also encompass things related to dogs like leashes and fire hydrants—may help to explain why the learned fear of a single object often blossoms into fear of many things. Finally, if the child’s fears are clinically problematic, he may receive an exposure-based therapy like systematic desensitization (see Chapter 80), which is thought to harness extinction of classical conditioning. Fear-eliminating benefits should then spread through the equivalence class (right box), even if most of the stimuli are not directly involved in the therapy.

Generativity

An experience is *generative* if it spawns novel abilities. The essential feature of stimulus equivalence is that many abilities emerge spontaneously after direct experience learning only a few component relations. In the example of Figure 39.1, two abilities were taught (relating A to B, and relating A to C). As many as seven others emerged, including two transitive relations (B➝C, C➝B), two symmetrical relations (B➝A, C➝A), and three reflexive relations (A➝A, B➝B, C➝C). Critically, these emergent abilities arise from experience with a properly structured “curriculum” that involves interconnected relations among stimuli. An individual who learns disconnected relations (e.g., “dog”➝Doberman photo and smiley-face icon➝printed word *happy*) can expect no emergent benefits.

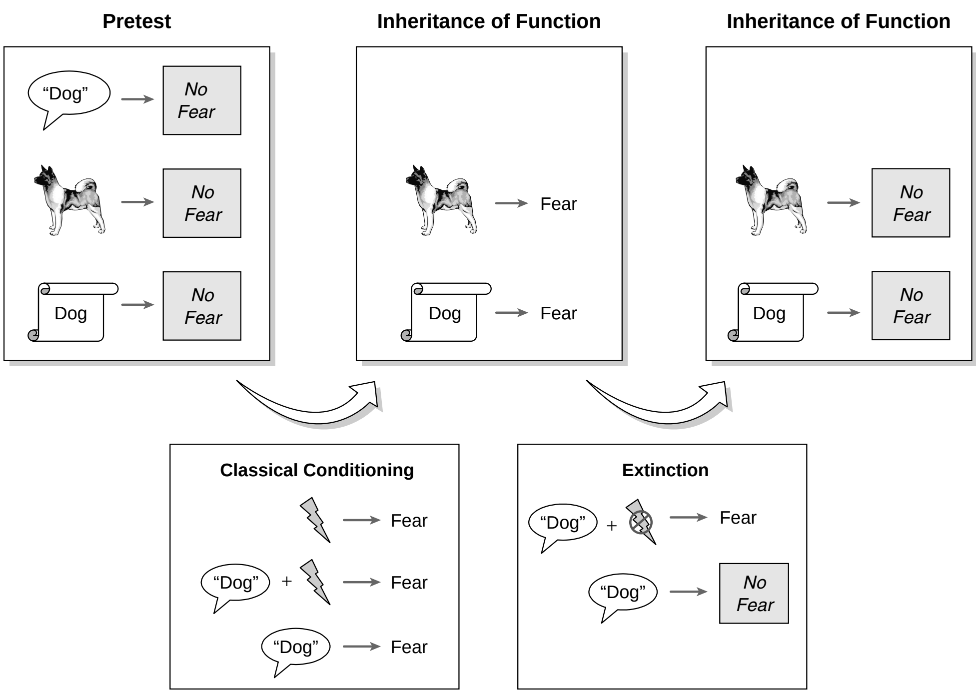

Figure 39.3 An experiment illustrating the inheritance of learned fear through an equivalence class. See the text for details.

SOURCE: Based on Dougher, 1998.

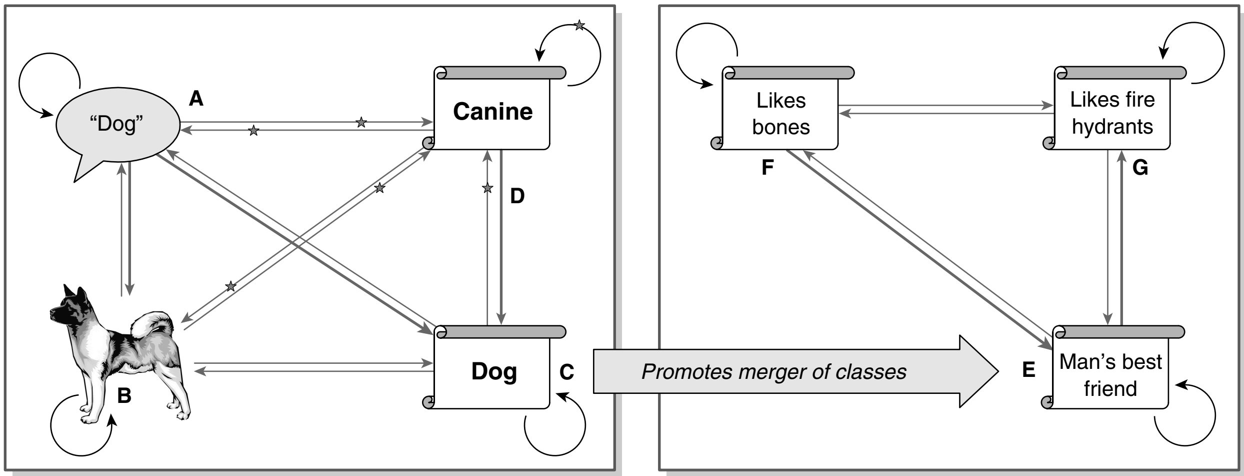

Figure 39.4 *Left box:* Each new stimulus added to a class greatly expands the number of emergent relations. *Both boxes:* Classes also can expand by merging with other classes. In both cases, see the text for explanation.

Boosting Class Size

The number of possible emergent abilities increases with the number of stimuli that are imbedded in interconnected relations. To illustrate, the left box of Figure 39.4 shows the implications for the person whose experience is summarized in Figure 39.1, that of learning that the printed word *canine* implies *dog* (for now, ignore the right box and the arrow connecting the two boxes). Learning this one ability spawns as many as six new emergent ones: four transitive, one symmetrical, and one reflexive (these relations are designated by stars in the left box). Several different approaches exist for building connections between relations (see Green & Saunders, 1998, for details). For many of these relations, the ratio of directly learned to acquired relations is N–1 to N^2, with N representing the number of stimuli in the class. Thus, in a three-member class, as per Figure 39.1, the ratio is 2 directly learned relations to 9 acquired (7 emergent). In a four-member class, as per Figure 39.4, the ratio is 3 to 16 (13 emergent). In a five-member class, the ratio is 4 to 25 (21 emergent), and so on.

Another factor that contributes to generativity is that two classes sharing a common member may merge into one bigger class. As illustrated in Figure 39.4 (right box), imagine learning that *man's best friend* ◊ *likes fire hydrants*, and *man's best friend* ◊ *likes bones*. This experience should create an equivalence class that includes the transitive relation between *likes bones* and *likes fire hydrants*. Now, as shown by the arrow spanning the two boxes, imagine that you subsequently learn that *dog* ◊ *man's best friend*. This experience places *man's best friend* into the *dog* class, and *dog* into the class containing *man's best friend*. As a result, every member of the *dog* class (A, B, C, D) should be readily associable with every member of the class containing *man's best friend* (E, F, G), and vice versa. Many new relations emerge between stimuli that have never been experienced together (such as "dog" ◊ *likes fire hydrants* and *likes bones* ◊ *canine*). In this case, teaching a single new relation yields a total of up to 22 new emergent transitive relations.[2]

CONNECTIONS

Stimulus Generalization and Cognitive Transfer

As will be described later, stimulus equivalence is thought to arise out of operant learning, but it is different from the operant learning concepts you probably read about in your introductory psychology textbook (see Chapter 36, Operant Conditioning). The purpose of this section is to clear up a common confusion about how stimulus equivalence is related to operant learning.

Brief Review of Operant Learning

Operant behavior is altered by its consequences, or the events that follow behavior. Among consequences,

[2] Note that class merger sometimes is constrained by a process known as contextual control, which allows a stimulus to belong to different classes under different circumstances, without those classes merging. Suppose you have a pet Doberman named Vlad. You decide to enter Vlad in a dog show, and when you fill out the registration forms, your dog may be equivalent to other Dobermans, but not to poodles, pugs, or Pekinese. When you select the supermarket aisle from which to obtain your pet's food, Vlad is equivalent to dogs of many breeds, but not to cats or parrots. When you discuss the responsibilities of pet ownership with a friend, Vlad is part of a class that includes many pet species, but does not include non-pet species like squirrels and coyotes, and not individual animals that aren't pets, such as a stray Doberman. Without contextual control, everything would become psychologically equivalent to everything else. See Sidman (1994) for a description of how contextual control arises.

reinforcers increase the probability that a given behavior will be repeated, and *punishers* decrease that probability. For example, a young student's first days in public school are full of these consequences. Some actions (e.g., sitting quietly and listening) are reinforced (e.g., by teacher praise) and over time they begin to increase in frequency. Others (e.g., eating chalk) are punished and begin to disappear.

Operant learning is situation-bound. What we learn through reinforcement and punishment tends to apply mainly to the situations in which the original learning occurred (Dinsmoor, 1995). Thus, a student can learn appropriate classroom behavior from consequences, but this skill set is expressed mainly while in school. Students do *not* suddenly whip out books and begin taking notes in a supermarket, restaurant, or automobile dealership. Operant behavior becomes situation-specific when the consequences that fuel it also are situation-specific. For example, academic behaviors are reinforced in some settings but not others. Behaviors like taking notes and sitting quietly with books have beneficial consequences in the classroom, but not in a restaurant. The term for a situation correlated with reinforcement of a given behavior is *discriminative stimulus*. By uniting the demands of the present moment with the benefits of past experience, a discriminative stimulus creates a situation-specific increase in the probability of acting a certain way (Dinsmoor, 1995). For academic behaviors, the classroom is a discriminative stimulus, but a restaurant is not.

Stimulus Generalization

Once discriminative stimuli exist, the benefits of past experience can spread to new situations through *stimulus generalization*. That is, operant learning spontaneously transfers to new situations that are perceptually similar to the discriminative stimulus (Dinsmoor, 1995). This process is how sitting quietly and taking notes can occur, without special training, the first time a student attends a new class or perhaps, when older, a professional conference.

Analogously, cognitive psychologists have long been interested in the study of transfer, which they define as the application of knowledge to novel contexts (Bransford, Brown, & Cocking, 2000). For instance, imagine that when quite young you learned to make change during a sale while working in your parents' business. You would show transfer if you could now also solve mathematics problems not involving money. This type of transfer shares two important characteristics with operant generalization: It involves the application of a learned ability to a new situation, and it is most likely to occur when there are close similarities between the original learning situation and the transfer context.

Different Processes

Stimulus generalization and cognitive transfer are special cases in which inheritance of function is mediated by *perceptual similarities* between the original learning situation and new stimuli. A conference hall, for instance, shares many features with a classroom (orderly rows of chairs, podium or desk at the front, etc.). When first encountered, it is "treated as the same" as a classroom because it is, in fact, similar to the classroom. By contrast, when first encountered a neighborhood bar is "treated differently" from a classroom because it is perceptually different from the classroom.

Importantly, stimulus equivalence is *not* stimulus generalization (transfer). Stimulus equivalence does *not* require that stimuli share perceptual features to be "treated as the same." Recalling Figure 39.1, the auditory waveforms of the spoken word "dog," the ink marks of the written word "dog," and the color patterns of the Doberman photograph would not be mistaken for one another, even by an alien visitor to our planet. They "go together" only insofar as experience (e.g., the arrows of Figure 39.1) teaches that they do.

A Productive Collaboration

Although stimulus equivalence and generalization (transfer) are different, these processes may combine to dramatically increase the size of equivalence classes (Fields & Reeve, 2000). Note, for instance, that although no two photographs of a Doberman are identical, most Doberman photographs look pretty much like any other. Doberman photographs also share perceptual features with some drawings of dogs, some real dogs, and some toy stuffed dogs. Practically speaking, in Figure 39.1 the B stimulus can represent not just a single Doberman photograph but a vast range of stimuli with shared properties. The equivalence class in Figure 39.1, therefore, incorporates not three stimuli but perhaps many hundreds or thousands. To be clear, overlapping stimulus relations place the first Doberman photograph into this class, whereas stimulus generalization then recruits many similar stimuli into it.

Semantic Generalization and Knowledge Networks

Stimulus equivalence may well remind you of other concepts you have encountered in your study of psychology because psychologists have long assumed that thinking is organized around networks of interconnected knowledge or ideas. Here are just three examples. First, the brain contains networks of synapses connecting neurons (see Chapter 3, Biological Psychology), which has prompted speculation that complex thought arises from these networks. Second, neural networks are the inspiration for computer programs that can simulate complex psychological abilities. The software building blocks of these programs are intended to mimic interconnected networks of neurons that fire together under certain circumstances (see Chapter 52, Artificial Intelligence). Third, some theorists believe that the contents of long-term memory are organized into *semantic networks* in which bits of knowledge are linked through shared meaning rather than physical similarity (see Chapter 46). All of these network theories share a common theme with stimulus equivalence: Individual parts of a knowledge network combine to make emergent abilities possible. In all of these cases, the whole of thought is seen as greater than the sum of its parts.

One interesting phenomenon that network theories anticipate is "semantic generalization," in which learning transfers to new situations on the basis of *shared meaning* rather than perceptual similarity. Consider a phobia patient with a crippling fear of spiders. Stimulus generalization may lead him to fear tomatoes because the cluster of slender leaves atop this fruit resembles a spider's leg. Later, through "semantic generalization," he may come to fear a variety of fruits and vegetables, but not because they remind him of spiders. Rather, through experience, these foods are understood similarly to tomatoes (for instance, they all come from plants, bear seeds, and make for healthy eating).

Knowing about stimulus equivalence can advance our understanding of "semantic generalization" in two ways. One advantage is that we can identify a contradiction in the term used to label this effect. Recall that, in "semantic generalization," transfer is mediated by shared meaning between the original learning situation and new stimuli. By definition, however, in generalization, transfer is mediated by perceptual similarity. Thus "semantic generalization" is not really *generalization*. Instead, it may be an example of the inheritance of function that we expect to see in equivalence classes (for instance, for most people who eat plant products, tomatoes probably are in an equivalence class with corn, carrots, and kumquats). Thinking of "semantic generalization" in this way avoids a confusion of terms. A second advantage comes from our ability to meet the "engineering challenge" mentioned earlier in this chapter. Some network theories do a better job of describing knowledge networks than of showing how to create them. By contrast, we know a great deal about how equivalence classes form, and this knowledge suggests a clear way in which the shared meaning underlying "semantic generalization" could arise through experience.

Concept Learning

Some psychologists believe that knowledge networks are created through *concept learning*. Speaking loosely, a concept is a collection of things that "belong together," and concept learning is the process of determining which things "belong together" (Barsalou, 1992). For example, a person who has acquired the concept of *chairs* not only recalls many specific chairs from past experiences but also recognizes new chairs, never before seen, for what they are. At the root of concept learning is experience, with feedback, of reacting to examples and nonexamples of the concept (similar experience helps to form equivalence classes). Imagine, for example, the experience of an alien visitor to our planet. This visitor might treat some objects as "chair" by trying to sit in them—sometimes with success (the object supports weight comfortably) and sometimes not (the object collapses or feels uncomfortable). The visitor might also react to some objects as "not chair" by continuing to stand—this decision works well in some cases, and in others the visitor is corrected by being told to "have a seat." Such experience creates two outcomes that are reminiscent of stimulus equivalence. First, various nonidentical examples and ideas coalesce into a kind of knowledge network. Second, a defining property of these knowledge networks is inheritance of function: When we learn something new about one example of a concept, our understanding of others automatically is updated.

There are at least two important differences between stimulus equivalence and concept learning. First, many concepts are defined by physical similarities among stimuli, while the stimuli in an equivalence class are united by shared function (Fields & Reeve, 2000; Zentall, Galizio, & Critchfield, 2002). Second, unlike in equivalence classes, the stimuli in some concepts are not interchangeable (Barsalou, 1992). They are united instead by other kinds of relations. For example, your favorite recliner is one of many chairs (equivalence relation), but it is also a component of your living room suite (part-to-whole relation). Psychologist Thomas Zentall, who has studied concept learning extensively, suggests that we think of equivalence classes as a special type of concept (Zentall et al., 2002).

METHODS

A good understanding of stimulus equivalence will require reading beyond this chapter including primary empirical reports. The procedures used to study stimulus equivalence in the laboratory are rather technical (Green & Saunders, 1998), and as a result, many people find them to be difficult reading. The goal of this section is to explain enough about the methods of stimulus equivalence research to ease the burden of students who seek to explore published reports. Remember, however, that experimental procedures do not *define* stimulus equivalence; rather, they allow scientists to view the development of equivalence classes more clearly than would be possible in everyday situations. The "father of stimulus equivalence research," Murray Sidman (1994), has stressed that outside of the laboratory equivalence classes form under very different circumstances than those seen in the laboratory.

Training, Testing, and Experimental Designs

Every stimulus equivalence experiment has at least two parts. During a training phase, overlapping stimulus relations (like those described by solid arrows in Figure 39.1) are explicitly taught, with feedback provided for correct and incorrect responses. Although the relations taught in many studies seem quite simple, sufficient practice is provided to guarantee that they are learned well. In published studies, you'll see that intelligent people often master these relations easily, but inexplicably, occasionally they require a surprising amount of practice. For example, in an experiment conducted by one of the authors of this chapter, a college-student participant needed 648 trials to learn three A‡B relations involving "simple" geometric stimuli. What is "simple," apparently, is in the eye of the beholder!

After training concludes, a testing phase follows in which untaught relations (analogous to those defined by dashed arrows in Figure 39.1) are assessed. Laboratory research studies often test for all possible reflexive, symmetrical, and transitive relations. Applied studies may focus only on transitive relations, as in the drug-names study (Zinn, 2002). Note that, although feedback is a critical part of training (see the section on "extrinsic" motivation, below), none is provided during the test, so that "emergent" relations can be shown to be truly untaught. Participants may, however, be confused by this sudden loss of feedback. Thus, in many studies, at the end of training, feedback is gradually eliminated from some trials to prepare participants for the change.

The results of a stimulus equivalence study may be presented at two levels of analysis. First is the general question of how many participants met the definition of stimulus equivalence by demonstrating all of the reflexive, symmetrical, and transitive relations. Second is the more specific question of how many of the participants mastered each potential emergent relation.

Some experiments, like the drug-names study (Zinn, 2002), use familiar group-comparison experimental designs involving many participants. Others, however, merely demonstrate that a particular set of training experiences is sufficient to create equivalence classes. Many of these studies focus on as few as four to six individuals.[3] which is likely to raise red flags for students who have been taught in statistics class to distrust small samples. The small *N* (see Chapter 10, Single-Subject Designs) is used for two reasons. First, when all individuals move seamlessly from random responding to complete mastery, there is little need for inferential statistics to verify the results. Second, the complex performances that participants learn require very detailed analysis, something that is easiest to describe for only a few individuals.

"Artificiality" of Laboratory Research

Stimulus equivalence has been the inspiration for a variety of attempts to teach socially significant abilities. In laboratory research, however, great emphasis is placed on understanding how new equivalence classes form, and the resulting study may not look much like the everyday world. The present section describes a few ways and reasons that laboratory research procedures may be different from the conditions in which equivalence classes form outside the laboratory.

Skill-Building Emphasis

Although people form equivalence classes in everyday experience, most research does not focus on analyzing these preexisting classes. One reason is that, because each individual has a unique history, such classes may be different for different people. For example, you might associate tequila with pleasant travels and social events while your friend associates it with sin and damnation. Another reason is that scientists have difficulty figuring out what everyday experiences, at some undetermined point in the past, created each individual's equivalence classes. Recalling the "engineering challenge" mentioned earlier, basic research usually seeks to understand *general rules* that govern how equivalence classes are formed rather than conditions that created a specific class for a specific person, and it does so by building new classes from scratch.

Unusual Stimuli

In any study of learning, it is important to verify that new abilities arise from experience in the study, rather than from a prior learning history. Thus, the stimuli used in basic research may be chosen because they are *not* meaningfully related to everyday experience. Instead of stimuli like those shown in Figures 39.1 and 39.2, many studies use stimuli that are unlikely to trigger strong associations prior to the start of an experiment (see Figure 39.5).

Structured Procedures

Because equivalence classes are defined by relations among stimuli, researchers use experimental procedures that precisely define these relations. Most typically, stimuli are presented in a *match-to-sample* procedure, one variant of which is illustrated in Figure 39.5. In the top leftmost panel, a *sample stimulus* is presented. The participant may be required to point to it, or to position a mouse cursor over it, to show that she is paying attention (top row, second panel), which reveals two or more *comparison stimuli* that are

[3] A note about subject selection: You may be surprised to discover that many seminal experiments in this area focused on individuals with developmental disabilities. Although "reflective thinking" is important to everyone, as Dewey (1933) stressed, researchers have seen three advantages in studying persons with intellectual challenges (Sidman, 1994). First, essential to any learning experiment, it is easy to identify information that they do not know. By contrast, it is harder to find a subject matter about which higher-functioning individuals know little. Second, because persons with disabilities form equivalence classes with difficulty, studying them will reveal only the most reliable precursors of class formation. If a set of learning experiences works for this population, imagine how well it should work with typically developing individuals. Third, there is an ethical bonus in studying people with disabilities because they have the most to gain personally from the learning experiences of the study.

physically unlike the sample (top row, third panel). The researcher has arbitrarily decided that one of these "goes with" the sample, and the participant's job is to learn from feedback about this correspondence. In Figure 39.5, an octagon comparison corresponds to the sample zeta (top row); a circle comparison corresponds to the sample phi (middle row); and a triangle comparison corresponds to the sample delta (bottom row). Thus, the procedure teaches, not a simple discrimination in which some stimuli are correct and others incorrect, but rather a *conditional discrimination* in which the correct choice depends on which sample is present.

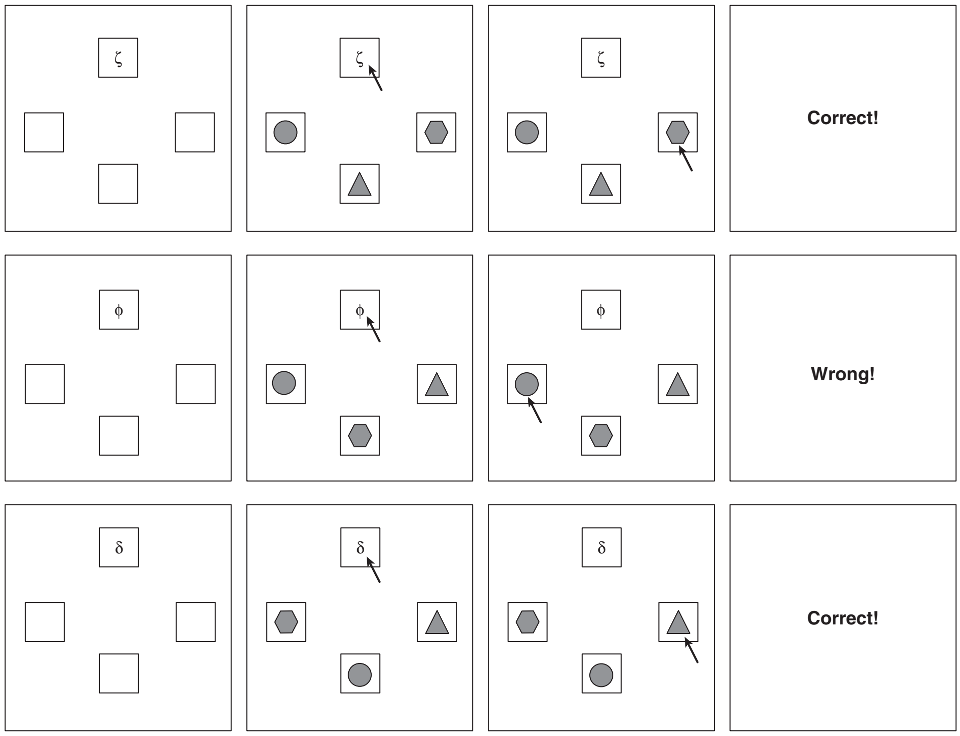

Figure 39.5 The match-to-sample procedure and examples of unfamiliar stimuli, often used in laboratory research on stimulus equivalence.

"Extrinsic" Motivation

If the correct stimulus is chosen, feedback indicates that the response is correct (Figure 39.5, top and bottom rows, last panel). Otherwise, corrective feedback follows (middle row, last panel). Feedback not only shapes the subject's skills; it also provides a form of motivation that is important to completing the experiment. Unlike many everyday stimuli, stimuli like those in Figure 39.5 are not inherently interesting. Because most people care about being correct, feedback gives participants a reason to pay attention and to learn. To enhance motivation further, some studies accompany feedback with money (e.g., two cents for each correct response) or other tangible benefits.

Multiple Classes

All experiments attempt to create two or more equivalence classes simultaneously. Each of the relations shown in Figure 39.5 (zeta-octagon, phi-circle, and delta-triangle) would be part of a different potential equivalence class (e.g., in A-B-C format, the classes might be: zeta-octagon-•, phi-circle-≈, and delta-triangle-≈). For economy of presentation, only the training of A ‡ B relations is shown. Creating multiple classes provides a built-in test of replicability—whatever works in creating one class should work in creating the others. Notice, too, that the comparison stimuli are the same in all parts of Figure 39.5. Training for multiple classes allows stimuli that are correct in one relation to be used as incorrect comparisons in others, which, in turn, eliminates some potential confounds that give the impression of participants understanding more than they really do (see Green & Saunders, 1998, for a nice explanation of this problem).

APPLICATIONS

Instruction

John Dewey, whose thoughts on "reflective thinking" began this chapter, also argued,

> The value of any fact or theory as bearing on human activity is, in the long run, determined by practical application—that is, by using it for accomplishing some definite purpose. If it works well—if it removes friction, frees activity, economizes effort, makes for richer results—it is valuable. (McLellan & Dewey, 1908, p. 195)

From this perspective, research on stimulus equivalence has shown considerable promise. Because basic research on stimulus equivalence has focused on building novel repertoires from the ground up, its procedures can, with relatively little modification, be employed to build socially relevant repertoires (Stromer, Mackay, & Stoddard, 1992). Applied stimulus equivalence studies always differ from their basic-research counterparts in one primary way: the stimuli are integral to real-world skills.

Because many academic topics require fluency with core equivalence relations, the range of possible stimuli is quite large. Sometimes stimulus equivalence is used to give an extra assist to people with intellectual challenges. In these cases, the relations worth teaching may be quite elementary (Figure 39.1 is a possible example). Sometimes stimulus equivalence is used for instructional efficiency in the teaching of typically developing children. Equivalence-based procedures have been used to teach important academic skills like the rudiments of a second language (e.g., equivalences between English and Spanish words) and of basic mathematics (e.g., equivalences between fractions, decimals, and the quantities they represent). Yet equivalence classes also are imbedded in more sophisticated knowledge. One study, for example, used equivalence-based instruction to teach college students the relations between factored and unfactored algebraic equations and the graphically portrayed functions that the equations represent (Ninness et al., 2005).

When thinking about the instructional potential of stimulus equivalence, it is important not to get too caught up in the detailed procedures of formal experiments. For instance, Zinn's (2002) drug-names study (see Figure 39.2) nicely illustrates how stimulus equivalence can inform the teaching of fairly complex information, but like all research, it focused as much on documenting an effect scientifically as on enhancing the lives of the participants. Murray Sidman (1994) has noted that although good instruction is carefully structured, it may omit many of the methodological bells and whistles seen in formal experiments. For example, training and testing may be intermingled, detailed explanations may complement the training experiences, emergent relations may be reinforced as soon as they occur, and reflexive and symmetrical relations may not be tested. As in laboratory studies, of course, the critical ingredient is a curriculum of overlapping stimulus relations. To get a sense of how learning experiences can be streamlined from the strict protocols of the laboratory, try teaching yourself a few facts on drugs and behavior by following the procedures described in Table 39.1.

Table 39.1 A streamlined approach to learning facts about psychoactive drugs through stimulus equivalence. Construct a deck of index cards containing the information shown below. Study them by looking at the front, and saying or writing what's on the back. Shuffle the deck each time you use it, and study until you can run through the deck without error. Then, take the test in the Appendix.

Front	*Back*
Stimulant	Pemoline
Stimulant	Cylert
Cylert	ADHD
Depressant	Alprazolam
Depressant	Xanax
Xanax	sleeplessness
Opiate	Fentanyl
Opiate	Sublimaze
Sublimaze	pain
Neuroleptic	Chlorpromazine
Neuroleptic	Thorazine
Chlorpromazine	Schizophrenia

Interpretations of Pressing Social Problems

Some stimulus equivalence experiments are intended to model processes that are thought to contribute to socially important problems. Recall earlier examples on the development of consumer brand-name preferences and the spread of learned fear. Other experiments have explored the possible role of stimulus equivalence in sexual arousal, fantasy, and deviance (Barnes & Roche, 1997); in how individuals understand social relations (Schusterman, C. R. Kastak, & D. Kastak, 2003); and in social stereotyping, as illustrated by the following research-based example (described by Dixon, Dymond, Rehfeldt, Roche, & Zlomke, 2003). Perhaps you are told that, on average, the people of a country we'll call Nation X drink more alcohol than people in many other countries (i.e., Nation X ‡ heavy drinkers). Experience also tells you that people who have overindulged in alcohol are not in peak intellectual form (heavy drinkers ◊ dimwitted). Through intuitive inference, you may have low intellectual expectations for people of Nation X (Nation X ◊ dimwitted). This false syllogism may be hard to resist, and illustrates how "reflective thinking" is not always beneficial.

An understanding of equivalence classes has helped theorists to speculate on the origins of many pressing problems like child abuse (Keenan, McGlinchey, Fairhurst, & Dillenberger, 2000), psychological disorders like depression (Dougher, 1998), and prejudice (Dixon et al, 2003). These interpretations are too involved to summarize here, but a brief example may help to illustrate what is involved. Dixon et al. proposed that emergent social stereotyping, similar to what was described in the last paragraph, can explain the hostile attitudes that many people in the United States hold, in the wake of the 2001 terrorist attacks, toward individuals of Middle Eastern descent. A similar process may underlie the hostile attitudes toward the United States of people in underdeveloped countries (including some who become terrorists). Dixon et al. believe that an equivalence-based analysis of these problems has two practical benefits. First, the analysis predicts that some popular educational strategies for reducing prejudice actually will magnify the problem. Research bears out this prediction. Second, the analysis suggests some novel strategies for reducing prejudice and, potentially, terrorism.

THEORY

Implications of Stimulus Equivalence for Theories About Behavior

Stimulus equivalence research arose out of behavioral psychology (see Chapter 2, Psychology in the 20th

Century), which is based on learning processes like operant learning. Psychologist Stephen Hayes, who has led the effort to apply stimulus equivalence concepts to clinical and everyday phenomena, has argued that stimulus equivalence forces two changes in how we look at behavioral psychology. First, behavioral psychology sometimes is criticized as too simple to account for complex human abilities such as language and thought (Barsalou, 1992). Because of its generative potential, however, stimulus equivalence may be one contributor to these abilities.

Second, stimulus equivalence may clarify some specific aspects of theories about behavior. Consider *conditioned reinforcers*, which become effective consequences through learning (see Chapter 36, Operant Conditioning). Traditionally, this type of learning is thought to be similar to what Russian physiologist Ivan Pavlov observed in his famous studies of conditioned reflexes in dogs. You may recall from your introductory psychology course that Pavlov paired an unimportant sound (tone) with a stimulus (meat) that elicited an unlearned reflex (salivation). With repeated pairings, the tone began to elicit salivation. Analogously, most textbooks will tell you that conditioned reinforcers arise when formerly unimportant stimuli are paired directly with stimuli that already work as reinforcers (e.g., see Dinsmoor, 1995). A smile, for instance, is reinforcing to most people, but not originally to infants. It may become reinforcing to the extent that infants encounter caretaker smiles at the same time as reinforcers such as food and warmth.

Inheritance of function provides a different way in which conditioned reinforcers can be created. For example, many people enjoy ice cream, suggesting that, for these people, ice cream can function as a reinforcer. Through inheritance of function, any stimulus that becomes associated with ice cream in an equivalence class also may become a reinforcer, even if the association is indirect. If you know your classical conditioning well, you may recognize this process as a possible recipe for the phenomenon called higher-order classical conditioning. Yet there are reasons to think that classical conditioning cannot fully explain conditioned reinforcement. Here are two: First, classical conditioning applies only to elicited responses such as reflexes, but not everyone would agree that "being reinforced" by a given stimulus involves a reflex. Second, higher-order classical conditioning often is weak, while conditioned reinforcement is a powerful part of the everyday world. Thinking of conditioned reinforcement as inheritance of function avoids both of these problems. Psychological effects other than classical conditioning are known to spread through equivalence classes, and the resulting effects can be quite potent.

Theories About Stimulus Equivalence

The conditions under which equivalence classes arise are becoming well understood, but theories of what equivalence is and why it occurs are still in development. Three of these emerging theories are now described briefly.

Murray Sidman (2000) proposed that stimulus equivalence is simply a complete explanation of operant learning. According to Sidman, any time operant learning occurs, all of its components parts—situational cues, responses, and the consequences that affect behavior—become members of an equivalence class. This straightforward view anticipates much of what we know about equivalence class formation in people, yet it is hard to reconcile with two types of findings. First, some studies suggest that equivalence classes sometimes form after experience that does not appear to involve operant learning, raising the question of whether stimulus equivalence is strictly an operant phenomenon. Second, it remains to be understood why nonhumans, who otherwise show operant learning quite readily, have difficulty forming equivalence classes (Zentall et al., 2002).

Two alternative views are based on the idea that equivalence-class formation is not synonymous with operant learning, but rather arises from certain kinds of operant learning histories. *Naming theory* states that stimulus equivalence becomes possible once learning (including operant learning) creates both receptive and expressive language skills (Horne & Lowe, 1996). According to naming theory, equivalence classes form when an individual assigns the same linguistic label (expressive repertoire) to stimuli that are not otherwise alike (the expressive language repertoire). Once assigned, this common label then prompts the individual to understand the stimuli as interchangeable in other ways (receptive language repertoire). Naming theory has some empirical support. Stimulus equivalence seems to begin in young children about the time that they become language capable, and people who initially fail to form an equivalence class may succeed after being taught a common name for the stimuli. One criticism is that naming theory is not really testable. For example, sometimes people appear to form equivalence classes without using common names for the stimuli. This result could mean that naming was absent (in which case naming theory is wrong). It could also mean that naming occurred but simply was not expressed aloud (in which case there is no objective way to verify that naming occurred).

Relational frame theory suggests that stimulus equivalence results from a skill that must be learned directly at first, and later spontaneously transfers to new circumstances (Hayes, D. Barnes-Holmes, & Roche, 2001). According to this view, early in life we learn which stimuli are interchangeable—one class of stimuli at a time, and only through the brute force of direct experience. There are no emergent relations. Thus, very young children experience the world much like the Complete Instruction Group in Zinn's (2002) study on teaching drug names. After many stimulus classes have been mastered in this inefficient way, a generalized ability emerges. We develop a sort of conceptual template for recognizing any set of interchangeable stimuli. Now equivalence classes are readily formed, complete with emergent relations, any time that environmental circumstances expose us to overlapping stimulus relations. Note that relational frame theory is broader than the other two theories. It assumes that equivalence is just one of many possible conceptual templates for relations among stimuli (others include "opposite from," "more than," and "part of"). A major criticism of relational frame theory, however, is that it is vague about the type and amount of prior experience that leads to the generalized ability described.

All of these theories have stimulated important new research, but there is no agreement about which of them is most accurate. What unites these theories is an attempt to relate stimulus equivalence to simpler learning processes. As suggested by our earlier discussion of network theories of knowledge, of course, it also may be possible to think about stimulus equivalence in terms of cognitive principles of memory and concept learning.

SUMMARY

Stimulus equivalence is one way in which individuals can "go beyond the information given" to understand things that have never been taught directly by experience. That this ability is important has never been controversial. That it arises predictably from certain kinds of experiences is only now being well understood. Research on stimulus equivalence creates the potential to engineer "reflective thinking" that is beneficial, and to understand situations in which "reflective thinking" can be counterproductive.

Much remains to be learned about stimulus equivalence, and contemporary research is unfolding on three major frontiers. First, if stimulus equivalence offers general insights into human understanding, then more studies are needed to explore its relevance to a wide range of everyday circumstances. For instance, where instruction is concerned, stimulus equivalence procedures should be useful in teaching about inferential statistics, biology, physics—any subject in which equivalences exist among terms and concepts. Also, it is important to move from merely describing pressing problems like terrorism in stimulus equivalence terms to using insights based on stimulus equivalence to address these problems in socially significant ways.

Second, the connections between stimulus equivalence and more traditional topics in psychology, such as concept learning, neural networks, and the structure of long-term memory, remain to be mapped out fully. So far, one can point to general similarities, but much research is needed to see which specific effects seen in stimulus equivalence investigations also appear in these other phenomena, and vice versa. Such research will provide a more coherent view of the intricacies of complex thinking.

Finally, and in some ways most intriguingly, research is exploring the extent to which stimulus equivalence is a uniquely human phenomenon. Where learning is concerned, nonhuman animals are surprisingly similar to us, exhibiting many of the same characteristics of operant conditioning, concept learning, and memory (e.g., Zentall et al., 2002). They do not, however, readily show stimulus equivalence. Students of psychology often wonder what it is that makes human beings special. Research on stimulus equivalence may help to answer this question.

Authors' Note

By commenting on a draft of this chapter, these Illinois State University students helped the author to see a difficult topic through student eyes: Jennifer Hitt, Katie Enright, Danielle Helmer, Sara Gambrel, Lena Gustafson, and Cassidy Powell.

REFERENCES AND FURTHER READINGS

Barnes, D., & Roche, B. (1997). Relational frame theory and the experimental analysis of human sexual arousal: Some interpretative implications. In K. Dillenberger, M. O'Reilly, & M. Keenan (Eds.), *Advances in behaviour analysis* (pp. 184–204). Dublin, UK: University College Press.

Barnes-Holmes, D., Keane, J., Barnes-Holmes, Y., & Smeets, P. M. (2000). A derived transfer of emotive functions as a means of establishing differential preferences for soft drinks. *The Psychological Record*, *50*, 493–511.

Barsalou, L.W. (1992). *Cognitive psychology: An overview for cognitive scientists.* Hillsdale, NJ: Erlbaum.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: D.C. Heath.

Dinsmoor, J. A. (1995). Stimulus control: I. *The Behavior Analyst, 18,* 51–68.

Dixon, M. R., Dymond, S., Rehfeldt, R. A., Roche, B, & Zlomke, K. R. (2003). Terrorism and relational frame theory. *Behavior and Social Issues, 12*, 129–147.

Dougher, M. J. (1998). Stimulus equivalence and the untrained acquisition of stimulus functions. *Behavior Therapy*, *29*, 577–591.

Fields, L., & Reeve, K. F. (2000). Synthesizing equivalence classes and natural categories from perceptual and relational classes. In J. C. Leslie & D. Blackman (Eds.), *Experimental and applied analysis of human behavior* (pp. 59–83). Reno, NV: Context Press.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). New York: Plenum Press.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.) (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic/Plenum.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 185–241.

Keenan, M., McGlinchey, A., Fairhurst, C., & Dillenburger, K. (2000). Accuracy of disclosure and contextual control in child abuse: Developing procedures within the stimulus equivalence paradigm. *Behavior and Social Issues, 10,* 1–17.

McLellan, J. A., & Dewey, J. (1908). *The psychology of number and its applications to methods of teaching arithmetic*. New York: D. Appleton & Co.

Ninness, C., Rumph, R., McCullen, G., Harrison, C., Ford, A. M., & Ninness, S. K. (2005). Functional analytic approach to computer-interactive mathematics. *Journal of Applied Behavior Analysis, 38*, 1–22.

Schusterman, R. J., Kastak, C. R., & Kastak, D. (2003). Equivalence classification as an approach to social knowledge: From sea lions to simians. In F. B. M. de Waal & P. L. Tyack (Eds.), *Animal social complexity: Intelligence, culture, and individualized societies* (pp. 179–206). Cambridge, MA: Harvard University Press.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127–146.

Stromer, R., Mackay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education, 2,* 225–256.

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*, 237–248.

Zinn, T. E. (2002). Using stimulus equivalence to teach drug names: A component analysis and drug name classification procedure. (Doctoral dissertation, Auburn University, 2002). *Dissertation Abstracts International, 63*, 3051.

APPENDIX

Self-Test on Drug Facts

Answer the questions in order. Once you have answered a question, do not return to it. Correct answers are shown at bottom.

1. Stimulants are prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

2. Neuroleptics are prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

3. Opiates are prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

4. Depressants are prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

5. Pemoline is also called
a. Xanax. c. Sublimaze.
b. Thorazine. d. Cylert.

6. Alprazolam is also called
a. Xanax. c. Sublimaze.
b. Thorazine. d. Cylert.

7. Chlorpromazine is also called
a. Xanax. c. Sublimaze.
b. Thorazine. d. Cylert.

8. Fentanyl is also called
a. Xanax. c. Sublimaze.
b. Thorazine. d. Cylert.

9. Pemoline is prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

10. Alprazolam is prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

11. Chlorpromazine is prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

12. Fentanyl is prescribed to relieve
a. pain. c. ADHD symptoms.
b. schizophrenia symptoms. d. sleeplessness.

Answer Key

All of the questions probed for transitive emergent relations. You did not directly study any of these facts. 1-C. 2-B. 3-A. 4-D. 5-D. 6-A. 7-B. 8-C. 9-C. 10-D. 11-B. 12-A.

PART VII

Individual Differences and Personality

40

Psychometrics

Marcel Satsky Kerr
Texas Wesleyan University

The field of psychometrics aims to measure psychological phenomena. Researchers disagree as to whether and how psychological measurement is accomplished. In his letter to the Grand Duchess Christina, Galileo Galilei (1610/1957) stated that the universe has its own language and set of characters. If we wish to understand our universe, we must understand that it “was written in the language of mathematics.” (p. 237). This chapter introduces the field of psychometrics.

DEFINITION

When we analyze the roots of the word *psychometrics,* we find *psycho,* which means “individual” or “mind,” and *metric,* which means “measurement.” The field of psychometrics attempts to measure psychological phenomena. A more specific working definition of psychometrics is derived from the field’s historical roots. Galileo’s philosophy about the importance of numbers is reiterated in the works of Sir Francis Galton. Galton (1879) defined *psychometry* as the “art of imposing measurement and number upon operations of the mind” (p. 149). Taken together, these philosophies suggest that psychometrics focuses its efforts in the quantification of psychological constructs. That is, as a discipline it assumes that all worthy human attributes can somehow be quantified. Today, this definition is applied more broadly to include more than mental operations. Modern psychometrics attempts to measure all aspects of the human condition. These include our *cognitions*—thoughts, beliefs, attitudes, and perceptions; our *behaviors*—overt (observable) and covert (within the individual), intentional and unintentional, and personal and social; and our *emotions*—positive, negative, primary, and secondary. Measuring these human attributes poses many challenges, as illustrated in the colorful history of psychometrics as a discipline.

HISTORY

Psychometrics has a home in the broader field of psychology, an arm of the social and behavioral sciences. To understand its role in this family of sciences, one must first understand its conception and birth. A brief history of psychology is needed. Whereas 18th-century philosophers (*Immanuel Kant, Erasmus Darwin,* and *Ernst Heinrich Weber*) planted the idea of psychology as a science, it took the works of 19th-century physiologists and physicists (*Gustav Theodor Fechner, Charles Darwin*, *Herbert Spencer,* and *Wilhelm Wundt*) to develop the seedling field into a full-fledged discipline with true experimental methodology. The thread that united these diverse thinkers was physics, which was the exemplar science they hoped psychology would become. Psychology’s scientific path was altered, however, by the onset of World War II (1939) and the events and policy decisions that occurred thereafter. Grave human atrocities were witnessed by Allied and

Axis powers alike. Many of the infringements upon human rights were done in the name of scientific advancement. With the world astonished, leading psychologists were well positioned, in both Germany and the United States, to effect social reform to a level unparalleled today.

In this manner, modern psychometrics was conceived by two parents—physics and bureaucracy. From its ancestry in physics, psychometrics gained objectivity, quantification, parametric statistics, systematic methodology, and inductive and deductive reasoning. From its ancestry in bureaucracy, psychometrics gained pragmatics, census and surveys, nonparametric statistics, qualitative methodology, administration, and reform. Evidence of the field's bipolar ancestry remains today, as mentors continue to train their apprentices in one approach or the other. Thus, emerging psychometric researchers tend to be skilled in either quantitative or qualitative methodology. Even though their different viewpoints are clear, contemporary psychometricians are more compatible than they are combative. Both approaches have much to contribute to our understanding of psychology.

THE IMPORTANCE OF PSYCHOMETRICS TO PSYCHOLOGY AND INDIVIDUAL DIFFERENCES

Like any true science, psychology is a dynamic discipline. Individuals develop, societies evolve, and variables of interest change. Therefore, our measurement tools and procedures must continually change as well. The use of quantitative and qualitative methods have produced thousands of research findings that have advanced our understanding of personality, neurology, memory, intelligence, learning, emotion, and child development. Likewise, these psychometric methods have advanced our knowledge in areas outside of psychology such as medicine, pharmacology, engineering, education, sociology, and politics. Over the decades, consistent bodies of findings emerge in each area, lending *reliability* (consistency over time or across samples) to our knowledge bases. To this end, certain areas have exhausted efforts aimed at understanding aggregate effects across individuals. These research outcomes spurred the field of *phenomenology* (the study of individual human perceptions).

Where objective, aggregate approaches leave off, phenomenological approaches promise addition to our knowledge bases by looking inward, into human consciousness. Phenomenological research employs qualitative methods of inquiry, as it is more interested in how individual differences determine behavior than in one's objective responses to stimuli. Consequently, phenomenologists rely on descriptive procedures rather than inferential. As history has witnessed often, the emergence of an integrated approach was inevitable. Thus, the field of *psychophysics* was born, which is the study of the relation between physical stimuli and the subjective perceptions they produce. Both the quantitative and qualitative approaches have much to offer the evolving science of measurement. The quantitative approach has contributed objective procedures that allow us to estimate differences across individuals, and the qualitative approach has contributed procedures such as content analysis that yields a rich picture of the individual as he sees the world.

THEORY AND MODELING

The vast majority of psychometricians have been trained in the quantitative tradition. Therefore, the majority of current psychometric procedures were derived from two sets of theories—*classical test theories* and *modern latent trait theories*. Classical test theories share a history that spans the centuries back to Sir Francis Galton, and they are referred to as "classic" because of their reliance on *true score theory. True score theory* posits that any measurement may be broken down into two parts—true scores and error scores. For example, if a metric measuring tape is used to measure a group of adults' heights, each individual's height consists of two portions: the individual's true height in centimeters and some degree of measurement error, hopefully a minimal few millimeters or fractions thereof. The primary goals of classical test theories (CTT) are to analyze individual differences across test scores and to improve the reliability of measurement tools. They do not aim to understand individual test scores. Therefore, results from studies rooted in CTT may be *generalized* (applied broadly) to populations but not to the level of the individual.

Modern latent trait theories, more commonly referred to as *item response theories,* share a history back to early 20th-century work on intelligence (*Alfred Binet*), but its recent popularity surged in the 1960s with item response models developed by *Georg Rasch*. Item response theories (IRT) are models that endeavor to relate person characteristics and item characteristics to the probability of a *discrete* (categorical) outcome such as "yes" or "no" answers to test questions. Similar to CTT, item response theory aims to improve reliability of measurement tools. However, IRT specializes in improving reliability through its investigation into the psychometric properties of assessments. Both CCT and IRT use a series of quantitative procedures to determine measurement reliability such as *correlation* (a number reflecting the linear relation among variables) and *covariance* (how much variables overlap in their variances). The key element that differentiates the two test theories is how *models* (visual representations of relations among variables) are used to assess a measure's psychometric properties. Whereas CCT strives to identify a model that best fits a set of data, IRT strives to identify or obtain data that fit a model. It is a simple case of *inductive* (from the specific to the general) versus *deductive* (from the general to the specific) reasoning, respectively.

The Model Fits the Data

With the goal of identifying the model that *best* fits the data, classical methods analyze *matrices* (rectangular arrays of numbers) of correlations and covariances. Several classical methods exist. The four most commonly used classical analyses include (a) *factor analysis*, modeling aimed at explaining variability among observed variables with a smaller set of unobserved factors; (b) *cluster analysis*, modeling that partitions a dataset into subsets with shared traits; (c) *path analysis*, a form of multiple regression that is causal modeling aimed at predicting *exogenous variables* (dependent variables) from *endogenous variables* (independent variables); and (d) *structural equation modeling*, a form of confirmatory factor analysis aimed at testing theory by modeling social behavioral constructs as latent variables.

Because each of the aforementioned methods uses datasets taken from a sample, results obtained from them may differ across samples. Another limitation of classical methods has been documented for over two decades by David Rogosa and colleagues. Rogosa (2004) asserts that because classical models assess *between-subjects* (across individuals) variability and assume that *within-subjects* (within individuals) variability has a constant rate of change, they inadequately measure individual differences in response to a *treatment effect* (intervention). For this reason he suggests that most classical test methods are misleading, as individual growth curves are not considered appropriately.

The Data Fit the Model

With the goal of obtaining data that fit a specified model, IRT statistical procedures vary dramatically from those employed by classical testing. The primary procedure of IRT modeling relates person characteristics and item characteristics to item responses, enabling the improvement of the assessment tool's reliability. IRT uses logistic models to estimate person parameters (e.g., reading comprehension level) and one or more item parameters (e.g., difficulty, discrimination). In this manner, IRT models have two advantages. First, unlike CTT, which only ascertains an assessment's average reliability, IRT models allow the careful shaping of reliability for different ranges of ability by including only the *best* items. Thus, IRT models yield stronger reliability findings. Second, because IRT models provide greater flexibility in situations when data are used at different times, with different samples, and with varied test forms, measures derived from IRT are not dependent upon the sample tested.

Although item response theories are fairly new and developing, there are two disadvantages that prevent them from having mainstream appeal. These disadvantages are related. First, IRT relies on complex logistical modeling, and many psychometric researchers lack the mathematical and procedural skills to employ and interpret them. Second, CTT analyses are readily available in user-friendly and readily available applications such as *SAS* (statistical analysis software) and *SPSS* (formerly, the Statistical Package for the Social Sciences). IRT modeling applications are rare, more expensive, and not currently available in SAS or SPSS. However, IRT methodology continues to develop, making the impact of these new modeling approaches promising.

METHODS

All psychometric methods are one of two types, qualitative or quantitative. *Qualitative* methods investigate the reasoning behind human behaviors, and therefore attempt to answer how and why humans behave as they do. *Quantitative* methods assume that all constructs investigated can be measured numerically, regardless of whether they are overt or covert. Thus, quantitative methods allow us to answer the what, where, and when of human behavior. For these reasons, qualitative measures are considered subjective tests and consist primarily of *open-ended* (free response) items, whereas quantitative measures aspire to be objective and consist mostly of *forced-choice* (must select from available options) items.

Regardless of method, the very first step in every psychometric project is to define the *target population* (those to which results may be generalized) and the *test sample* (representative subset of the population). Next, the psychometrician must decide upon operational definitions of the constructs investigated. *Operational definitions* describe constructs by the specific manner in which they are measured. It is the operational definitions that determine all subsequent steps in the psychometric process. Let's consider a psychometric project with the goal of *scale construction* (measurement tool development). The researchers wish to develop a tool that measures teenagers' understanding and use of the Internet. The next step in a qualitative approach would be to write open-ended questions or items such as "For what purposes do you use the Internet?" or "Why is using the Internet helpful to you?" In using a quantitative approach, there is one more important consideration before measurement items are written. The researchers need to determine the variables' scale of measurement.

Scales of Measurement

A variable's level or *scale of measurement* refers to the nature of the values assigned to describe it accurately. Because of its simplicity and ease of understanding, the research field has adopted the classification system developed by Stanley Smith Stevens (1946). Stevens's scales of measurement consist of four levels, organized from simplest to most sophisticated. Therefore, variables may be scaled at the *nominal, ordinal, interval,* or *ratio level*. A variable's scale of measure is important, as it determines how items will be written, how many items are needed, the data collection method, the adequate sample size needed, and ultimately what statistical analyses may be used to determine the tool's psychometric properties. Each of the four scales of measurement is discussed in turn.

Nominal

Nominal variables are categorical. The values of the variable consist of names or labels and are mutually exclusive. The nominal scale is not continuous and merely reflects membership in a group. Examples of nominal variables are sex (female or male), political party affiliation (Democratic, Republican, Independent), and favorite ice cream (chocolate, strawberry, vanilla). The only comparisons that can be made between nominal variables are statements of equality or inequality. No mathematical operations are appropriate. In scale construction projects, many demographic questions regarding the sample are written at the nominal level. For example, "What is your sex?" or "Select your current grade in school." In such cases, participants are forced to select either "female" or "male" and their grade level from a set list of options (i.e., 9th grade, 10th grade, 11th grade, or 12th grade). Only descriptive analyses such as *frequencies* (how often each value appears in the variable) and percentages are appropriate for assessing nominal variables.

Ordinal

Ordinal variables contain all the characteristics of nominal variables and also contain a rank order of the constructs measured. For this reason, the values of ordinal variables are called *ordinals*. The rank-order characteristic of these variables makes them more sophisticated than nominal variables, as *greater than* and *less than* comparisons can be made. Like the nominal scale, the ordinal scale is not truly continuous; therefore, higher mathematical operations are not appropriate. Examples of ordinal variables include marathon finalists (1st place, 2nd place, 3rd place), university classifications (freshman, sophomore, junior, senior), and social economic status (lower, middle, upper). At a glance, ordinal variables look just like nominal categories; however, inherent within the values is a rank order. For example, university classification is more accurately depicted as ordinal because seniors have more completed credits than juniors, who have more completed credits than sophomores. In a scale construction project, researchers use ordinal variables to assess demographic information as well as attitudes or preferences. For example, "On a scale of one to five, how much do you prefer to shop online compared to shopping in a store?" Like nominal variables, ordinal variables are assessed by descriptive analyses such as frequencies and percentages, and the most typical values in the variable may be identified by two measures of central tendency, the *median* (the middle-most value) or the *mode* (most frequently appearing value).

Interval

The interval scale contains the characteristics of the nominal and ordinal scales as well as the assumption that the distances between *anchors* (values) are equivalent. These equal intervals allow meaningful comparisons among the variable's values, and mathematical operations such as addition and subtraction are appropriate. The researcher assigns anchors that may consist of positive or negative values. If zero is used, the point it represents is arbitrary. That is, zero does not reflect the true absence or degree of the construct measured. Rather, zero is a meaningful placeholder. For example, Celsius temperature is an interval scale, as zero degrees reflects the point at which water freezes; it does not reflect the point at which temperature ceases to exist. Other interval scaled variables include standardized intelligence tests (IQ) and the SAT (Scholastic Aptitude Test). The interval scale is useful, as ratios of differences across individuals may be expressed; however, higher operations such as multiplication and division are still not appropriate. For this reason, they are not truly numeric, although many researchers choose to treat their interval measurements as such. The interval scale's level of sophistication allows variables to be analyzed by numerous descriptive statistics such as the mode, median, and *mean* (arithmetic average), as well as *standard deviation* (average distance of values to the mean).

Ratio

The most sophisticated level of measurement is ratio. Ratio-scaled variables are continuous and truly numeric. The ratio scale contains all the characteristics of the prior three in addition to having a true zero point. That is, zero reflects the absence of the variable measured. Examples of ratio-scaled variables are time, money, length, and weight. Ratio statements are appropriate and meaningful, as it is accurate to conclude that a person who took 1.5 hours to complete her final exam was twice as fast as the person who completed the exam in 3 hours. All mathematic operations may be used, and the full range of descriptive statistics and *inferential statistics* (use of representative samples to estimate population parameters) may be employed for assessment of ratio variables.

Measurement Level Considerations

Sometimes the scale of variables is unclear. In such cases, the investigator defines the scale of measurement as she generates the variable's operational definition. For example, in measuring attitudes or preferences, some consider the scale ordinal, and others, interval. Likewise, many rating scales that meet the characteristics of the interval level are treated as truly numeric, thus ratio, by the investigator. These ambiguities may occur when a Likert scale is used. *Likert scales* are usually considered ordinal and use numbers as anchors, each of which reflects a level of agreement or disagreement with the statement at hand. The scale was developed by Rensis Likert (1932), an educator and industrial-organizational psychologist, who used the scale to evaluate management styles. It is the most commonly used scale in survey research. In scoring a survey, many researchers sum responses across items. When this procedure is employed, the resulting sum or mean may be *normally distributed* (have a bell-shaped curve), allowing the new score to be treated as an interval variable.

Psychometrics in Scale Construction

Regardless of whether the measurement instrument is an interview, self-report survey, checklist, or standardized assessment, measurement is an integral part of the scale construction process. For these purposes, *scale* is defined as a collection of items to which participants respond in a meaningful way. The responses are scored per their level of measurement and combined to yield *scale scores* (aggregate values used for comparison). Scale scores are then evaluated to determine how accurately and reliably the instrument measured the construct under investigation. Thus, the scale construction process unfolds in three stages: *scale design, scale development*, and *scale evaluation*.

Scale Design

As mentioned, a clear theoretical definition of the construct at hand must first be in place. Next, in any scale construction project, certain assumptions are made. The scale developer assumes that all participants can read at the same level and interpret the items similarly. Additionally, the developer assumes minimal differences in responses due to uncontrollable individual characteristics such as motivation, personality traits, and *self-preservation bias* (tendency to give socially desirable responses). The remaining steps in the scale design stage include (a) identifying and selecting the respondents; (b) determining the conditions in which the scale is administered; (c) identifying the appropriate analyses for scoring, scaling, and evaluating; (d) selecting content and writing items; and (e) sequencing the items.

Identifying and selecting respondents. A part of the operational definition of the construct depends on for whom the scale is intended. Identifying the target population is first. It seems to be a straightforward process, but without a clear population defined, to whom the results can be generalized becomes ambiguous and therefore subject to reader interpretation. For the use and understanding of the Internet project, one may define the target population as all teenagers (ages 13 to 19 years) in the United States who use the Internet. That is fairly specific, but there are still some gray areas. How often must they have used the Internet? At least once? For what purposes must they have used the Internet? Why limit it to U.S. teenagers? How will the participants be selected?

To answer the last question, several methods are possible, but the goal of each is the same—to acquire a sample of individuals who accurately represent the target population from which they were drawn. The ideal method to acquire such a sample is through *random sampling,* a process that assures that each and every member of the population has an equal chance of getting selected into the sample. An example of true random sampling includes placing all population member names into a hat and randomly selecting individuals until the desired sample size is obtained. Although it is the ideal, random sampling is rarely used in social science research, as it is often hard to find individuals who exhibit the constructs under investigation. For this reason, nonprobability sampling techniques are used, such as *quota sampling*, obtaining participants who reflect the numerical composition of subgroups on a trait in the population (e.g., depressed versus nondepressed individuals) and *haphazard sampling,* obtaining participants where they are conveniently available. Other questions may still emerge when one discerns how the scale will be administered.

How the scale is administered. The context and conditions in which the scale is administered have a profound effect on the scale's design. Scales may be administered individually or in groups. The responses may be *self-report* (individual's perception of self) or *other report* (individual's perception of another individual) accounts. Scales that measure ability may need to be timed and have very specific administration instructions. Scales may be administered orally, such as with interviews; in writing, such as with questionnaires and standardized tests; or physically, such as with *performance tasks* (e.g., finger tapping or puzzles). Finally, scales may be administered via different media. Examples include by telephone, mail system, or computer (i.e., e-mail or Internet). A careful consideration of media is recommended, as the choice affects one's *response rate* (percentage of completed scales returned) and ultimately the validity and reliability of the information gathered.

Selecting analyses. It is not premature to consider potential analyses in the scale design stage, as the steps in the scale construction process affect one another. If the developer does not know which analyses to employ to score, scale, and evaluate the information gathered, all previous efforts will have reduced value. The developer needs to identify the process, either by hand or by using a statistical software package that she will use to score the items. Scoring may involve reversing the direction of some items and deciding how to extract meaning from the responses. Choices include summing or averaging items to produce scale scores or *subscale scores* (aggregate values across a subset of items). Scale scores are usually assessed with a statistical package such as SAS or SPSS. To that end, statistical analyses must be selected to assess the items' and subscale scores' structure, validity, and reliability. This decision is also affected by the purpose and the design used to collect the data.

Selecting content and writing items. It is best to restrict a scale to a single construct, and it is advantageous to start by searching the literature for measures with a similar purpose. If a valid and reliable measure already exists, it begs the question as to whether a new one is needed. Next, the type of responses one wishes to capture depends on what is being measured. Do items ask for behavioral, cognitive, or affective responses to the construct? To address this question, many issues warrant consideration. Here we discuss five major issues.

First, *item format* is important. Items may be open-ended or *closed* (forced-choice). Open-ended questions are used when all possible answers are unknown, when the range of possible answers is very large, when the developer wishes to avoid suggesting answers to respondents, or when the developer wants answers in the respondents' own words. Closed questions are advantageous when there is a large number of respondents or questions, when administration time is limited, when scoring will be conducted by a computer, or when responses are to be compared across time or groups (B. Sommer & R. Sommer, 1991).

Second, the *response choices* depend on the underlying measurement dimension. For example, in rating statements, are respondents asked the degree to which they agree or disagree with them or the level of importance the rating statements have to them personally?

Third, to determine the *sensitivity* of items, the number of scale points is important. Fewer scale points (three to four) are less sensitive but increase the likelihood that respondents interpret the *anchors* (scale-point labels) similarly. Consequently, more scale points (seven to nine) are more sensitive but may reduce accurate anchor comprehension, thus increasing error. Regardless of number selected, the anchors should not overlap and should be inclusive, allowing an accurate choice for all respondents. Using an "other" category accomplishes the latter.

A fourth, related consideration is the *number* of items. The goal is to be thorough yet concise. Too few items will leave the construct under evaluated and less understood, although too many items may fatigue or bore respondents.

The fifth consideration in writing items is *wording*. The language used should be clear and meaningful. Items need to be written at the respondents' reading level, but jargon and catch phrases should be avoided. Likewise, the developer wants to avoid *double-barreled* (asking about more than one topic), *loaded* (emotionally charged), or negatively worded questions, as they increase the likelihood that respondents interpret items differently and ultimately skew the results.

Sequencing the items. Once items are written, the developer needs to consider the order of their presentation. If the items have different purposes, then more general items and those used to establish rapport should be listed first. More specific, more personal, and more controversial items should be listed later. Items need to flow logically from one to the other, ensuring that answers to earlier items do not influence the answering of later items. If there is a mixture of dimensions measured, then factual and behavioral items should be asked first, with perceptual and attitudinal items asked subsequently. If *reactivity* (completing items changes respondent's behavior) and/or self-presentation bias are unavoidable, then the presentation of items may be *counterbalanced* (presented in different order for some respondents) so that their possible effects may be determined in the scale-evaluation stage.

Scale Development

Now that the initial version of the scale has been produced, the next steps are taken to acquire preliminary findings that are used to edit the initial version until an acceptable final version of the scale is reached. Therefore, the scale-development stage is an iterative process that may use several different sample sizes and types. The first step may involve a *pilot test* (trial run with a small number of individuals). The goals of the pilot are to pretest the items for wording and clarity and to determine how easily instructions are followed and how long it takes to complete the scale. If the problems identified in the pilot test are minor, the scale is easily revised. If numerous problems are identified, a second pilot test may be warranted after the changes have been made.

With errors in respondent comprehension reduced, the second step in the scale-development stage is to administer the revised measure to a large sample of respondents who are representative of the target population. The goals of this second step depend upon the scale's purpose. For example, if the scale is constructed to identify the best set of items to measure teenagers' use and understanding of the Internet, then a data-reduction technique such as exploratory factor analysis may be employed. In *exploratory factor analysis*, scale items are interrelated, producing a correlation matrix. From these covariances an initial factor structure is extracted. Then the structure may be rotated to improve interpretability. Items with the highest factor loadings are retained and items with lower factor loadings and/or load on multiple factors are removed from the scale. In this manner, exploratory factor analysis is just one technique used to identify the fewest items that best measure an underlying construct.

Scale Evaluation

Once a stable factor structure has been established, measurement instruments are evaluated upon two criteria: *reliability*, the degree to which responses to the scale items are consistent, and *validity*, the degree to which the scale measures what it was developed to measure. In scale construction, reliability is determined by assessing the proportion of scale score variance that is not *error variance* (measurement noise). When constructing scales, many forms of validity may be determined. The most appropriate ones depend upon the purpose of the scale. For example, if the developer intends the new scale to be used as a diagnostic tool, then establishing both construct and criterion validity is important. *Construct validity* is the degree to which a measurement instrument accurately measures the theoretical construct it is designed to measure. In scale construction, construct validity is captured in the proportion of scale score variance that accurately represents the theoretical construct. *Criterion validity* (also referred to as *predictive validity*) is the degree to which a measurement instrument accurately predicts behavior on a criterion measure, where a *criterion measure* is any variable that the construct should be related to in a predictable manner. Therefore, in scale construction, criterion validity is represented by the proportion of criterion variance that is predicted by the scale. These forms of reliability and validity may be assessed by several different techniques. Specific techniques are discussed in the next two sections, as validity and reliability are important concepts beyond the task of scale construction; they are the primary goals of all psychometric endeavors.

Measuring and Improving Validity

Unlike research in the physical sciences, the constructs of interest in social and behavioral research are seldom physical, observable, or tangible phenomena. For example, psychologists want to understand concepts such as memory, prejudice, love, and job satisfaction, none of which is overt or contained in a Petri dish. For these reasons, social scientists have invested interest in measuring and ensuring validity. Over the years, determining validity has become an ongoing process that includes several steps and activities. As mentioned previously, the first step is important: It is a clear definition of the construct under investigation that includes how it is measured. Regardless of purpose, the social and behavioral researcher is interested in determining construct validity. Without this validity, the construct essentially does not exist. The strategies used to determine construct validity require one to evaluate the correlations between the construct measured and variables it is known to relate to theoretically in a meaningful way (Campbell & Fiske, 1959). Correlations that fit the expected pattern of relations provide evidence of the nature of the construct. In this manner, the construct validity, and thus value, of a theoretical construct is created through conclusions based upon an accumulation of correlations from various studies and samples in which the construct is measured. To this end, there are two primary forms of construct validity, convergent and discriminant.

Convergent validity is the degree to which the construct is positively related to other measures of the same construct or similar ones. For example, if one wished to develop a new measure of math ability for elementary schoolchildren, the new measure could be administered to a large sample of schoolchildren along with a test that required them to solve grade-level-appropriate math problems. The measure's convergent validity is evaluated by examining the correlation between the children's scale score on the new math test and their performance on the problem-solving task. These correlations are usually referred to as *validity coefficients.* Positive correlations (i.e., commonly .40–.60) suggest high construct validity, but there is no set criteria as to what range constitutes "adequate" construct validity; it depends greatly on the research purpose and discipline (Cronbach, 1971; Kaplan & Saccuzzo, 2005).

Discriminant validity is defined as the degree to which the construct is not related to constructs for which there is no theoretical basis that a relation exists. Said differently, it is the degree of divergence between the construct and the variables it should not be related to systematically. For example, to determine discriminant validity of the new math ability test, one might administer it to school-age children along with measures of reading ability and writing ability. Again, correlations among the math, reading, and writing scale scores are evaluated. Although they may be positively related (i.e., one's reading is related to interpreting math story problems), these validity coefficients should be low, indicating that the three constructs measured are distinct phenomena.

Taken together, indices of convergent and divergent validity provide social researchers with confidence that the constructs investigated in fact measure what they purport to and, over time, confirm their value to all who use them. Similar to scale construction, evaluating and improving validity are ongoing processes. Accordingly, determining measurement validity is the most important psychometric property of all scales, followed by determining the appropriate measures of reliability.

Measuring and Improving Reliability

At its most basic level, reliability reflects a measure's consistency or stability. It is not the case that a scale is reliable in all manners. Reliability may be assessed over time, with different age groups, in different contexts, or with different types of people. Similar to measures of validity, the appropriate measures of reliability depend upon the purpose of the research and nature of the construct. Here we discuss three reliability estimate procedures, with situations when they may be used.

First, *trait scales* are those that measure a construct considered relatively stable over time, such as intelligence and personality types. To evaluate trait scales one may wish to determine its *test-retest reliability* (consistency over time). To do so, one administers the measure to the same group of individuals at two different times. The scores across time should be similar. A correlation coefficient is computed between the time 1 and time 2 scale scores. Higher positive scores suggest good test-retest reliability. Importantly, how the trait is defined is a key element. When personality is considered a stable set of traits, an objective assessment such as the MMPI (Minnesota Multiphasic Personality Inventory) is used. When personality is considered less stable, akin to mood, a projective assessment such as the Rorschach Inkblot test may be used. In the latter case, test-retest reliability may be low and inappropriate.

Second, *state scales* are those that measure constructs that are expected to change in direction or intensity under different conditions. Our states (e.g., mood, motivation) are affected by many factors. To investigate state-scale reliability, one must answer one more question. Is the scale *homogenous* (measures a single construct) or *heterogeneous* (measures several related constructs). To assess homogenous scales, a measure of internal reliability is appropriate. *Internal reliability* (also referred to as *internal consistency*) is the degree of variability that exists among the scale items. When the variability is low, suggesting that all items measure the same construct consistently, internal reliability indices are high. When variability is high, internal reliability is reduced.

There are two procedures for assessing internal reliability: the split-half method or Cronbach's coefficient alpha. The *split-half* method consists of dividing the scale items into two equivalent sections. One might separate the odd- and even-numbered items or randomly place all items into two groups. Regardless, the same group of individuals completes both halves of the items, and their scores on the first half are compared to scores on the second half with a correlation. If indeed only one construct is measured, scores on both halves should be similar and the correlation coefficient should be positive and high, reflecting strong internal reliability.

When scales are homogenous, have no right or wrong answers, or are heterogeneous, Cronbach's coefficient alpha may be computed to estimate internal reliability. *Cronbach's alpha* measures variability among all scale items or those within a subscale. To illustrate, most standard IQ tests measure *general intelligence* (innate, nonspecific mental ability) as well as *specific intelligence* (developed domain-specific mental abilities) via subscales. For instance, there may be subscales that measure reading comprehension, verbal ability, problem solving, and logical reasoning. In this case, an alpha coefficient may be computed across all general items to yield an internal reliability estimate of general intelligence or computed among the subscale items to estimate how consistently the problem-solving items relate to one another. Cronbach alphas need to be in the .70 or higher range to be considered acceptable, although opinions on this criterion vary quite a bit (Kaplan & Saccuzzo, 2005).

As is true of determining validity, the process of ensuring the reliability of a scale is multifaceted and ongoing. The goal for both is continual improvement through the identification and reduction of error variance among the scale and related construct items. Once the field finds a scale's estimates of validity and reliability acceptable, its mainstream adoption and use are usually forthcoming.

APPLICATIONS

Where Psychometrics Is Practiced

You will find psychometricians working virtually anywhere—in the fields of psychology, counseling, government, marketing, forensics, education, medicine, and the military. Anywhere that research is conducted, psychometric experts are likely behind the scenes. In addition to the dimensions discussed previously (cognitions, behaviors, affects), field researchers are also interested in measuring individuals' *KSAs*, the specific Knowledge, Skills, and Abilities related to work in their particular job or profession. There is no terminal degree required in order to use the title *psychometrician*, although most hold a graduate degree (master's or doctorate) from a university. Most of those with formal graduate degrees received their psychometric training in an educational measurement program or quantitative psychology program. Consequently, specializations in experimental psychology are common among practicing psychometricians. For those wondering if a career in psychometrics might be a good personal fit, graduate internships in educational testing, healthcare consulting, and clinical applications are readily available.

Measuring Individual Differences

Much of the work conducted by the aforementioned professionals consists of assessing differences across individuals. Specifically, *individual difference psychologists* focus their examination on the ways individual people differ in their behaviors. Topics of investigation by these differential psychologists include intelligence, personality, motivation, self-esteem, and attitudes. For example, those interested in measuring intelligence are interested in understanding the differences in cognitive ability based upon age, sex, ethnicity, and other personal characteristics. It is simply not enough to gather scores from a large sample of diverse individuals, compute their average intelligence score, and generalize it back to the population with any meaning. For this reason, individual difference researchers believe that considering individual characteristics in our investigations is the way to contribute meaningfully to our knowledge bases. This rationale is why *norms* (means and standard deviations) for major standardized intelligence tests such as the *Stanford-Binet Scale* and *Wechsler Intelligence Scales* are available for specific age groups. Readers are encouraged to review Chapter 44, Intelligence, for more information on the theory and measurement of individual intelligence testing.

Group Testing

Individual test administration is a costly business, as a trained test administrator is required for each participant assessed; the time, effort, and resource costs accrue exponentially. Therefore, a large percentage of ability measurement now occurs in groups. In *group testing*, a trained examiner may read the test instructions to all participants at once and impose a time limit for the test. Participants read their own test items and record their scores, usually in writing. Participant responses usually consist of objective answers to forced-choice items, so that scoring and analysis are easily completed by one individual using a statistical software package. Clearly, the primary advantage of group testing is cost-effectiveness. Group testing occurs in schools at every level, in the military, and in industry, and researchers use group administration extensively for a variety of reasons. In these fields, group testing is used for various purposes including diagnostic screening, selection purposes, assessing special abilities, and assessing interest and aptitude levels for special jobs and occupational duties. Through group testing, measurement application is especially broad.

SUMMARY

Because of the prevalence of their use among psychometricians, the latter half of this chapter focused on quantitative methods and applications as opposed to qualitative. Qualitative methodologists continue to develop and fine-tune robust qualitative means of measurement. New developments emerge each year. For more information on qualitative methodology, readers should see Chapter 11, Qualitative Research.

Another area that promises psychometric advancement is item response theory modeling. Given the current trend of ever-increasing diversity in our society, models that aim to take into account individual growth curves are invaluable for an accurate understanding of human social behavior.

A final note for future directions is not novel. It is the importance of using multiple measurements in every research endeavor. The late *Donald T. Campbell*, considered by many the father of social research methodology, coauthored with *Donald W. Fiske* the paper titled "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." In it, Campbell and Fiske contend that single quantitative and qualitative methods have strengths and weaknesses that, individually, leave investigations lacking. They also warn investigators of problems that may arise when a single measurement of a construct is used for diagnostic decision making. They suggest that even the most stable of traits is variable under certain conditions. Half a century later, their message remains contemporarily important. This is why their 1959 paper is one of the most-often-cited articles in the social science research literature.

In this vein, convergent methodologies have emerged that center upon Campbell and Fiske's multimethod/multitrait principle of establishing validity. The term *triangulation* has been coined to describe how both quantitative and qualitative methods are blended to measure constructs more effectively (Webb, Campbell, Schwartz, & Sechrest, 1966). In such efforts, our faith in the validity of social constructs is maximized, as methodological error variance is reduced, if not altogether eliminated. To this end, psychometricians can have confidence that the constructs we measure truly mirror the underlying human qualities we endeavor to understand.

REFERENCES AND FURTHER READINGS

Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care, 42,* 1–16.

Campbell, D. T., & Fiske, D.W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cozby, P. C. (2001). *Methods in behavioral research* (7th ed.). Mountain View, CA: Mayfield Publishing.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Dawis, R. V. (1987). Scale construction. *Journal of Counseling Psychology, 34*(4), 481–489.

Galilei, G. (1957). *Discoveries and opinions of Galileo* (S. Drake, Trans.). New York: Anchor Books. (Original work published 1610)

Galton, F. (1879). Psychometric experiments. *Brain: A Journal of neurology, II,* 149–162.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*(4), 602–611.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140,* 44–53.

Mitchell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin, 3,* 398–407.

Porter, T. M. (2003). Measurement, objectivity, and trust. *Measurement, 1*(4), 214–255.

Professional Testing Inc. (2006). *What is a psychometrician?* Retrieved December 21, 2006, from http://www.proftesting.com/test_topics/pdfs/psychometrician.pdf

Rogosa, D. R. (2004). Some history on modeling the processes that generate the data. *Measurement: Interdisciplinary Research and Perspectives, 2,* 231–234.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 4*, 321–333.

Sommer, B., & Sommer, R. (1991). *A practical guide to behavioral research: Tools and techniques.* New York: Oxford University Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 278–286.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures*. Chicago: Rand McNally.

41

Testing and Assessment

John Juve
JurySync Litigation Consulting

The endeavor of measuring an intangible construct such as an individual's intelligence or aptitude is both fascinating and essential. Psychologists, educators, the government, and corporate America use the principles and measures of testing and assessment to solve challenging problems on a daily basis. For example, a psychologist may administer a wide variety of test batteries including intelligence and personality tests as well as measures to screen for neurological impairment. Educators may use tests to assess achievement and facilitate student placement decisions. Corporate America, as well as government agencies, have also embraced testing and frequently administer tests to make vocational decisions related to hiring, firing, and general utilization of personnel.

The field of testing and assessment in the United States is relatively young and consequently most of the major developments occurred during the 20th century. However, the origins of testing and assessment are neither recent nor American. Evidence has suggested that the ancient Chinese had a relatively sophisticated civil service testing program. For example, written exams were introduced by the Han Dynasty (202 BCE–200 CE) to measure an individual's potential to succeed in vocations such as civil law, military affairs, and agriculture.

Testing and assessment methods became quite well developed by the Ming Dynasty (1368–1644 CE) during which officials used a national multistage testing program to make vocational decisions in both local and regional venues. Individuals who performed well on local tests progressed to provincial capitals for more extensive examinations. For example, one relevant vocational measure included good penmanship, which was essential for clear and precise communication and therefore was viewed as a relevant predictor of suitability for civil service employment.

In the early 1800s, reports from British missionaries may have encouraged the English East India Company to copy the Chinese civil service system as a method of selecting employees for overseas duty. Because the testing programs worked well for the company, the British government eventually adopted and refined a similar system of testing for its civil service program. The French and German governments followed suit. In the late 1800s, the U.S. government established the American Civil Service Commission, which developed and administered competitive examinations for specified government jobs. The testing and assessment movement in the Western world grew rapidly from this point.

Most historians trace the early development of psychological testing to the investigation of individual differences that flourished in Europe (most notably Great Britain and Germany) in the late 1800s. There is no doubt that early experimentalists like Charles Darwin, Wilhelm Wundt, Francis Galton, and James McKeen Cattell laid an indelible foundation for 20th-century testing and assessment. Researchers believe that Darwin contributed one of the most basic concepts underlying psychological and educational measurement—individual differences.

Darwin's (1859) *On the Origin of Species by Means of Natural Selection* argued that chance variation in species would be perpetuated or encumbered based on its ability to adapt or survive in nature and consequently that humans had descended from the apes as a result of such chance genetic variation. Popular topics in contemporary psychology that reveal a noticeably strong Darwinian influence include theories of learning, developmental psychology, animal behavior, psychobiology, theories of emotions, behavioral genetics, abnormal psychology, and testing and assessment.

Darwin's research stimulated interest in the study of individual differences and demonstrated that studying human and animal behavior was at least as important as was studying the mind. Darwin's work appears to have influenced and motivated Francis Galton's research on heredity. Through his efforts to explore and quantify individual differences between people, Galton became a prominent contributor to the field of testing and measurement. Galton aspired to classify people according to their deviation from the average and eventually would be credited with contributing to the development of many contemporary psychological measures such as questionnaires, rating scales, and self-report inventories.

Galton and his assistant Karl Pearson pioneered the use of the correction coefficient. This development was important for the field of testing and measurement because it gave researchers a method for obtaining an index of a relation between two variables. Through Galton's research efforts and his persistence for educational institutions to maintain anthropometric records (e.g., height, breathing capacity, and discrimination of color) on their students, he encouraged widespread interest in the measurement of psychologically related topics.

Charles Spearman, who espoused a theory of intelligence in which he believed there was a single, global mental ability (general intelligence or "g"), also contributed significantly to the development of testing and measurement. Despite challenges to the concept, Spearman's "g" still permeates psychological thinking and research. Spearman has also been attributed with discovering that independent measures of an individual's physical characteristics (e.g., mental ability) vary in a random fashion from one measurement trial to another. In statistical terms, the correlation between such independent measures for a group of persons is not perfect. Because of this latter research, Spearman is often referred to as the father of classical reliability theory.

Wilhelm Wundt employed the early principles and methods of testing and measurement at his experimental psychology lab in Germany. For example, he and his students worked to formulate a general description of human abilities with respect to variables such as reaction time, perception, and attention span. Wundt believed that reaction time could supplement introspection as a technique for studying the elements and activities of the mind. Further, Wundt attempted to standardize his research methods and control extraneous variables in an effort to minimize measurement error. It appears he was ahead of his time in two respects. First, he attempted to control extraneous variables for the purpose of minimizing error, which is now a routine component of contemporary quantitative measurement. Second, he standardized research conditions, which is also a contemporary quantitative method used to ensure that differences in scores are the result of true differences among individuals.

Alfred Binet's pioneering work, which that led to the development of the first widely used intelligence test in 1905, also made an indelible impact on educational and psychological measurement. The French Ministry of Education commissioned Binet and Théophile Simon to devise a practical means to distinguish normal children from those with mental deficiencies. Binet concentrated his efforts on finding a way to measure higher mental processes and eventually devised a simple chronological age scale to determine a child's level of mental functioning.

The inception of the *Journal of Educational Psychology* in 1910 and its publication of Edward Lee Thorndike's seminal article, "The Contribution of Psychology to Education," were invaluable for the field of testing and measurement because they introduced and encouraged a common research forum to discuss educational measurement and testing issues. The journal's strict adherence to an experimental pedagogy had a significant impact on the credibility of research conducted in the field of educational psychology for the next 100 years. Thorndike and the other researchers played a vital role in establishing the fundamental theories and methods that would act to perpetuate and solidify the emergence of a separate field of psychological testing and assessment.

THEORY

Psychological testing and assessment have grown and matured primarily within the parameters of two fundamental theories—classical test theory and modern test theory. Both theories rely heavily on the essential concepts of reliability and validity to guide and substantiate credible testing and measurement practices.

Classical Test Theory

In the early part of the 20th century, researchers utilizing psychometrics (the science of psychological and educational measurement) focused primarily on the concepts of true score and measurement error, which were predominately based on Charles Spearman's correlation and reliability studies. These concepts were further strengthened by Lee Cronbach's discussion of construct validity in 1955. Research has indicated that classical test theory has matured due to several remarkable achievements over the past 150 years, including (a) a recognition of the presence of error in measurement, (b) a conception of error as a random variable, and (c) a conception of correlation and the means for it to be indexed.

In 1904 Charles Spearman demonstrated how to correct a correlation coefficient for attenuation due to measurement error and how to obtain the index needed to make the correction. According to Traub (1997), Spearman's demonstration marked the beginning of classical test theory. Additionally, Frederic Kuder, Marion Richardson, Louis Guttman, Melvin Nowak, and Frederic Lord contributed important ideas. For example, the Frederic Kuder and Marion Richardson internal consistency formulas (KR_{20} and KR_{21}) were published in 1937, and in 1945, Louis Guttman published an article titled "A Basis for Analyzing Test Reliability," in which the lower bounds of reliability were explicitly derived (Traub, 1997). The culmination of these efforts to formulize classical test theory was realized best by the research of M. R. Novick and Frederic Lord in the 1960s.

Measurement Error

Classical measurement theory, commonly referred to as traditional or true score theory, has provided a strong foundation for psychometric methods since its origin. In classical test theory, an assumption is made that each examinee has a true score on a test that would be obtained if it were not for random measurement error (Cohen & Swerdlik, 2002). A true score is considered to be measurement without error or, more simply, reducing the discrepancy between an observed score (with error) and the true score on a given measure.

The principles of psychometrics can be used as an effective tool to reduce error in mental measures. Psychometricians (testing and measurement professionals) should understand they cannot completely eliminate error in mental measures, and their goal should therefore entail finding methods to reduce the known sources of error in as many testing contexts as possible. Error associated with the measurement process can be described as any factor not directly relevant to the construct or topic being measured. Although there are many ways of categorizing the types of error in test scores, Lyman (1978) presented one useful classification system, in which test score errors were related to the following five factors: (a) the influence of time, (b) test content, (c) the test examiner or scorer, (d) the testing situation, and (e) the examinee.

Osterlind (2005) emphasizes that mental measurements are imperfect and attributes this to two human fallibilities. The first fallability is that humans do not always respond to a test item or exercise in a manner that typically reflects their best ability. For example, a sick or fatigued examinee may obtain a score that inaccurately reflects his or her true score. Similarly, individuals who are not motivated to do well on a test (or worse, purposely make mistakes or provide inaccurate information) can contribute error to their test scores.

The second human fallibility that may result in measurement error is our inability to produce a flawless testing instrument. Instruments may be imperfect because a test developer may not design an assessment with the precision needed for sound measurement. A testing instrument may also be imperfect because an item writer may create items or exercises that do not accurately represent the cognitive processes involved in measuring a specified task or construct. Further, instruments created to measure mental processes typically assess only a small component of a complex phenomenon. Consequently, we must infer meaning from a limited sample to a broader domain and this inferential process is not without error.

Random and Systematic Measurement Error

According to Nunnally and Bernstein (1994), it is common and appropriate to think of an obtained measure as deviating from a true value, and the resultant measurement error can incorporate a mixture of both systematic and random processes. When error is systematic, it can affect all observations equally and be a constant error or affect certain types of observations differently and consequently demonstrate bias. For example, a miscalibrated thermometer that always reads five degrees too low illustrates a constant error in the physical sciences.

It is important to understand that random error applies to an individual response, whereas systematic error applies to a group's response. More specifically, random error is the difference between a true score and an observed score for an individual, whereas systematic error comprises consistent differences between groups that are unrelated to the construct or skill being assessed. Random errors are considered much more common than systematic errors and are important because they limit the degree of lawfulness in nature by complicating relations (Osterlind, 2005).

Nunnally and Bernstein (1994) suggested that random error may influence scores on a particular classroom test by (a) the content studied (e.g., luck in studying the same information that is on the test), (b) luck in guessing, (c) state of alertness, (d) clerical errors, (e) distractions, (f) not giving your best effort, and (g) anxiety. Random measurement error can never be completely eliminated, but researchers should always strive to minimize the sources of this error. Essentially anything that detracts examinees from exhibiting their optimal score can be considered random error; keep in mind that random error always degrades the assessment of an individual's true score.

The determination of systematic error is often difficult to identify because it frequently relies on arbitrary and erratic judgments. According to Osterlind (2005), systematic error is generally associated with, but not limited to, differential performance on an exam by samples of one sex or a particular ethnic group. It can apply to any distinct subpopulation of an examinee group. Sometimes systematic error is referred to as test bias in older terminology (more recently differential performance), which oversimplifies a complex measurement phenomenon. Osterlind explained that it is the consistency in error that makes it systematic. For example, a compass can provide an inaccurate reading because of influences unrelated to its intended purpose (error), and the compass can give the inaccurate reading every time it is used (systematic).

Shortcomings of Classical Test Theory

Perhaps the most important shortcoming of classical test theory is that examinee and test characteristics cannot be separated. More specifically, the examinee and test characteristics can only be interpreted within the context of the other. The examinee characteristics of interest usually pertain to the proficiency measured by the test. According to Hambleton, Swaminathan, and Rogers (1991), in classical test theory the notion of proficiency is expressed by the true score or the expected value of an observed performance on the test of interest. In classical test theory, an examinee's proficiency can only be defined in terms of a particular test. The difficulty of a test can be defined as the proportion of examinees in a group of interest who answer the item correctly (Hambleton, Swaminathan, & Rogers, 1991). Whether an item is difficult or easy depends on the proficiency of the examinees being measured, and the proficiency of the examinee depends on whether the test items are difficult or easy. In other words, a test item or exercise can only be defined in terms of a reference group. Due to the limitations associated with classical test theory, measurement professionals have sought out alternative modern test theory methods to address the problems inherent with classical test theory.

Modern Test Theory

How has classical test theory influenced modern test theory? What are the differences between classical and modern test theory? How will testing and assessment professionals benefit from modern test theory beyond the limitations inherent within classical test theory? Considering and reflecting on these questions has challenged measurement professionals to examine their research methodology practices, data analysis procedures, and decision-making processes. The notion of a true score dates back to the time of early measurement theorists like Galton, Binet, and especially Spearman in the early 20th century. As one may anticipate, error is a feature that distinguishes various psychometric models or theories. Each measurement model defines error differently, and each approaches error from a distinct perspective.

Classical measurement theory is often referred to as a true score theory because of its emphasis on true scores, whereas many modern measurement theories (e.g., latent trait and generalizability) are referred to as theories of reliability or occasionally universe score theories (Osterlind, 2005). Nunnally and Bernstein (1994) suggested that contrasting what is classical test theory versus what is modern test theory is always a bit risky, but they consider measures based on linear combinations as classical. For example, Thurstone's law of comparative judgment is generally regarded as classical because it appeared more than 60 years ago, but it is not based upon linear combinations. Conversely, the 1950 Guttman scale is considered modern, despite its long history, because it is based upon individual response profiles rather than sums.

Although classical measurement theory has served many practical testing problems well for more than a century with little change, Osterlind (2005) believes that classical measurement theory is not comprehensive enough to address all theoretical and practical test problems. For example, in classical test theory, random error is presumed to be equal throughout the entire range of the score distribution; this assumption is not realistic. Although the standard error of measurement is extremely useful, it does not reveal differing error rates at various points in the distribution. Moreover, classical measurement theory treats each item and exercise on a test as equally as difficult as all other items and exercises; this assumption also is not realistic. Although classical test theory has its deficiencies, it does have the advantage of being simpler to understand and is more accessible to a wider measurement audience. Whether measurement professionals choose to utilize classical test theory, modern test theory, or both to address measurement issues, they will necessarily rely on foundational concepts such as reliability and validity to determine the consistency and meaningfulness of a test score or what the test score truly means.

Reliability

Reliability can be defined as the extent to which measurements are consistent or repeatable over time. Further, reliability may also be viewed as the extent to which measurements differ from occasion to occasion as a function of measurement error. Reliability is seldom an all-or-nothing matter, as there are different types and degrees of reliability. A reliability coefficient is an index of reliability or, more specifically, a proportion that indicates the ratio between the true score variance on a test and the total variance (Cohen & Swerdlik, 2002).

Reliability has become an essential concept providing researchers with theoretical guidelines and mathematical measures to evaluate the quality of a psychological or educational construct. Stated differently, reliability measures help researchers identify and diminish the amount of error involved in measuring a psychological or educational construct. Error implies that there will always be some inaccuracy in our measurements, and measurement error is common in all fields of science. Psychological and educational specialists, however, have devoted a great deal of time and study to measurement error and its effects. More specifically, they have sought to identify the source and magnitude of such error and to develop methods by which it can be minimized. Generally, tests that are relatively free of measurement error are considered reliable, and tests containing measurement error are considered unreliable.

In testing and assessment settings, many factors complicate the measurement process because researchers are rarely interested in measuring simple concrete qualities such as height or length. Instead, researchers typically seek to measure complex and abstract traits such as intelligence or aptitude. Consequently, educational or psychological researchers must carefully assess the reliability and meaningfulness of their measurement tools (e.g., tests or questionnaires) in order to make better predictions or inferences regarding the phenomena they are studying.

When evaluating the reliability of a measure, researchers should first specify the source of measurement error they are trying to evaluate. If researchers are concerned about errors resulting from a test being administered at different times, they might consider employing a test-retest evaluation method, in which test scores obtained at two different points in time are correlated. On other occasions, researchers may be concerned about errors that arise because they have selected a small sample of items to represent a larger conceptualized domain. To evaluate this type of measurement error, researchers could use a method that assesses the internal consistency of the test, such as the split-half evaluation method.

Although reliability is a critical factor in determining the value of a test or assessment, it is not a sufficient condition in and of itself. Testing and measurement professionals must also evaluate the validity of a test or assessment.

Validity

The term *validity* may engender different interpretations based on the purpose of its use (e.g., everyday language or legal terminology). However, when the term is used to describe a test, validity typically refers to a judgment pertaining to how effectively the test measures what it purports to measure. More specifically, the term is used to express a judgment based on acquired evidence regarding the appropriateness of the inferences drawn from test scores (Cohen & Swerdlik, 2002). The most recent standards for educational and psychological testing, published in 1999, emphasize that validity is a unitary concept representing a compilation of evidence supporting the intended interpretation of a measure. Some commonly accepted forms of evidence that researchers may use to support the unitary concept of validity are (a) content validity, (b) criterion-related validity, and (c) construct validity.

Content validity evidence has typically been of greatest concern to educational testing and may be described as a judgment concerning the adequacy with which a test measures behavior that is representative of the universe of behavior it was designed to measure. *Criterion-related validity* evidence may be described as evidence demonstrating that a test score corresponds to an accurate measure of interest. Finally, *construct validity* may be described as a judgment related to the appropriateness of inferences drawn from test scores regarding individual standings on a variable referred to as a construct (*e.g.*, intelligence).

METHODS AND APPLICATIONS

Researchers can use classical and modern test theory methods and applications to develop, maintain, and revise tests or assessments intended to measure academic achievement, intelligence, and aptitude or potential to succeed in a specific academic or employment setting. The assessment of aptitude and achievement began some time after the assessment of intelligence and was aimed primarily at identifying more specific abilities. Intelligence tests (e.g., the Stanford-Binet and the Wechsler) were useful because they produced valuable assessment information about overall intellectual level (global intelligence), but limited because they yielded little information about special abilities. The development of aptitude and achievement tests was an attempt to bridge this gap (Walsh & Betz, 2001). Aptitude tests were thought to measure people’s ability to learn if given the opportunity (future performance), whereas achievement tests were thought to measure what people had in fact learned (present performance).

Measuring Academic Achievement

Measuring academic achievement can be a daunting and challenging task that requires constructing test items that accurately measure student learning objectives. The process of constructing good test items is a hybrid of art and science. Typically, test creators who seek to measure academic achievement are interested in developing selected response (e.g., multiple-choice) and supply-response (e.g., essay) test items. However, before developing such a test, test developers must intimately understand the important relation between classroom instruction and the subsequent assessment methods and goals. A necessary condition for effective and meaningful instruction involves the coexistence and continuous development of the instructional, learning, and assessment processes.

The relation of instruction and assessment becomes evident when instructors closely examine the roles of each process. For example, Gronlund (2003) emphasized this relation when he stated, “Instruction is most effective when directed toward a clearly defined set of intended learning outcomes and assessment is most effective when designed to assess a clearly defined set of intended learning outcomes” (p. 4). Essentially, the roles of instruction and assessment are inseparable.

Pedagogic research (e.g., Ory & Ryan, 1993) has encouraged instructors to use Bloom’s taxonomy to analyze the compatibility of their instructional process, their desired student outcomes or objectives, and their test items. Instructors typically use test scores to make inferences about student content mastery. Consequently, it is essential that these inferences be valid. The inferences made by instructors are more likely to be valid when the test items are comprised of a representative sample of course content, objectives, and difficulty. Therefore, when developing test items, instructors should revisit the relation between course objectives, instruction, and testing. The assessment process involves more than merely constructing test items. Instructors should also become adept at administering, scoring, and interpreting objective and essay type tests.

Multiple-Choice Test Items

Test items are typically presented as objective (multiple-choice, true and false, and matching) and essay type items. Multiple-choice items represent the most frequently used selected-response format in college classrooms (Jacobs & Chase, 1992), and instructors should become adept at creating and revising them. The benefits of using multiple-choice items include (a) accurate and efficient scoring, (b) improved score reliability, (c) a wide sampling of learning objectives, and (d) the ability to obtain diagnostic information from incorrect answers. The limitations include (a) the increased time necessary for creating items that accurately discriminate mastery from nonmastery performance, (b) the difficulty of creating items that measure complex learning objectives (e.g., items that measure the ability to synthesize information), (c) the difficulty and time-consuming nature of creating plausible distracters (i.e., incorrect response alternatives), (d) the increased potential for students to benefit from guessing, and (e) the possibility that the difficulty of the items becomes a function of reading ability, even though reading ability may not be the purpose of the assessment.

True-False Test Items

True-false item formats typically measure the ability to determine whether declarative statements are correct. Effective true-false items are difficult to construct because they usually reflect isolated statements with no (or limited) frame of reference (Thorndike, 1997). The benefits of using true-false items include (a) ease of construction (when compared with multiple-choice items), (b) accurate and efficient scoring, (c) flexibility in measuring learning objectives, and (d) the usefulness for measuring outcomes or objectives with two possible alternatives. The limitations include, (a) an increased guessing potential, (b) the difficulty of creating unequivocally true or false items, (c) a measurement of typically trivial knowledge, and (d) a lack of diagnostic information from incorrect responses (unless students are required to change false statements into true statements).

Matching Test Items

Matching items typically measure associative learning or simple recall, but they can assess more complex learning objectives (Jacobs & Chase, 1992). The benefits of using matching items include (a) ease of construction (b) accurate and efficient scoring and (c) short reading and response times. The limitations include (a) measurement of simple recall and associations (b) difficulty in selecting homogenous or similar sets of stimuli and response choices and (c) provision of unintended clues for response choices.

Essay Test Items

Essay questions are more useful than selection-type items when measuring the ability to organize, integrate, and express ideas (Gronlund, 2003). The benefits of using essay items include (a) ease of construction, (b) measurement of more complex learning objectives, and (c) effective measurement of the ability to organize, compose, and logically express relations or ideas. The limitations include: (a) time consumption for grading, (b) decreased score reliability when compared with objective items, (c) limited ability to sample several learning objectives due to time constraints, and (d) the items' typical dependence on language.

Measuring Intelligence

The measurement of intelligence is perhaps one of the most controversial topics that testing and assessment professionals encounter. To begin, there is a great deal of debate regarding a common definition of what entails intelligence and how it can best be measured. Intelligence may be defined as a multifaceted capacity that manifests itself in different ways across the developmental lifespan, but in general it includes the abilities and capabilities to acquire and apply knowledge, to reason logically, to plan effectively, to infer perceptively, to exhibit sound judgment and problem-solving ability, to grasp and visualize concepts, to be mentally alert and intuitive, to be able to find the right words and thoughts with facility, and to be able to cope, adjust, and make the most of new situations (Cohen & Swerdlik, 2002). Although this definition appears to be comprehensive, it also demonstrates the inherent difficulty involved with measuring an individual's intelligence given the broad range of factors that can potentially be associated with intelligence.

Two common contemporary measures of intelligence include the Stanford-Binet Intelligence Scale and the Wechsler Tests. The fourth edition of the Stanford-Binet contains 15 separate subtests yielding scores in the following four areas of cognitive ability: verbal reasoning, abstract/visual reasoning, quantitative reasoning, and short-term memory (Cohen & Swerdlik, 2002). The Weschler tests were designed to assess the intellectual abilities of people ranging in age from preschool (ages 3 to 7), childhood (ages 6 to 16), and adulthood (ages 16 to 89). The tests are similar in structure, and each contains several verbal and performance scales.

Measuring Aptitude or Ability

Classical and modern test theory methods and applications can effectively be used to develop, maintain, and revise tests or assessments intended to measure aptitude or ability. For example, when researchers or admissions committees want to know which students should be selected for a graduate program (e.g., medical or law school), they often depend on aptitude measures to predict future behavior or inclinations. The forecasting function of a test is actually a type or form of criterion validity known as *predictive validity evidence* (Kaplan & Saccuzzo, 2001). For example, a student's score on the MCAT (medical college admissions test) may serve as predictive validity evidence if it accurately predicts how well that particular student will perform in medical school. The purpose of the test is to predict the student's likelihood of succeeding on the criterion—that is, successfully achieving or meeting the academic requirements set forth by the medical school. A valid test for this purpose would help admissions committees to make better decisions because it would provide evidence as to which students would typically succeed in an academic medical school setting.

Measuring Occupational Aptitude

Business corporations and the government typically use aptitude tests to facilitate their decision-making processes concerning employment recruitment, placement, and promotion. An example of an occupational assessment is the General Aptitude Test Battery (GATB), which is a reading ability test that purportedly measures aptitude for a variety of occupations. The U.S. Employment Service developed the GATB to help make employment decisions in government agencies. The GATB seeks to measure a wide array of aptitudes, ranging from general intelligence to manual dexterity (Kaplan & Saccuzzo, 2001). The Department of Defense utilizes the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB yields scores that apply to both educational and military settings. In the latter, the ASVAB results are used to facilitate the identification of students who may qualify for entry into the military, and they can potentially be used by military officials to recommend the assignment of soldiers to various occupational training programs (Kaplan & Saccuzzo, 2001).

FUTURE DIRECTIONS

Attempting to predict the future trends of educational or psychological testing and assessment can be beneficial for several reasons. The examination of trends can inform current practice, clarify future research goals, and identify areas of potential concern or danger. Before attempting to speculate on the future of testing and assessment, it seems prudent to contemplate several historical questions raised by Engelhard (1997). For example, what is the history of educational and social science measurement? Have we made progress, and if so, how should progress be defined within the context of measurement theory and practice? Who are the major measurement theorists in the social sciences, and what are their contributions? What are the major measurement problems in education and the social sciences, and how have our views and approaches to these problems (i.e., reliability, validity, test bias, and objectivity) changed over time?

Reflecting upon Engelhard's questions is critical for measurement professionals in the 21st century; they surely will be confronted with an unprecedented array of social, legal, economic, technical, ethical, and educational issues. More important, measurement professionals will likely have the opportunity to influence the roles of localization, teacher education, computerized testing, assessment variation, and litigation pertaining to the field of testing and assessment by becoming an integral part of future policy debate.

Localization

Cizek's (1993) thought-provoking article, "Some Thoughts on Educational Testing: Measurement Policy Issues into the Next Millennium," discusses localization or a trend for elementary and secondary education programs to develop, administer, and interpret tests and assessments to be utilized at the district, school, or classroom level. Localization represents a departure from reliance on national and commercially produced tests and assessments.

The trend to limit the use of national testing will inevitably place more responsibility and accountability on school districts, schools, and teachers to become more actively involved in the assessment process. Localization proponents believe it is valuable for teachers to be intimately involved in constructing, administering, and interpreting assessments relevant to their students' educational needs. However, some measurement professionals have expressed concern that most teachers do not receive adequate formal training in testing and assessment, unless they pursue a master's degree.

Rudman (1987) has discussed an alternative to the full localization approach; he believes that the future of assessment will witness a solicitation on the part of some test publishers to offer tailored instruments to both local and state agencies. These tailored instruments would likely contain items from previous editions of standardized survey-type achievement tests and national test-item banks. These items would be combined with the so-called local items supplied by the classroom teachers to provide the ability to compare the local items with the previously standardized items to examine the appropriateness of the relevant psychometric properties.

Computer-Assisted Testing

The convenience and economy of time in administering, scoring, and interpreting tests afforded by computer-assisted testing and measurement will continue to evolve and play an important role in the future of testing and assessment professionals. Computer-assisted assessment is unlike conventional testing. where all examinees receive all items. Rather, its focus is on providing each examinee with a unique assessment based on an examinee's proficiency level. When this objective is accomplished, the result is known as a *tailored* or *adaptive test*. According to Nunnally and Bernstein (1994), if testing is under computer control, the tailored test is referred to as a *computerized adaptive test* (CAT). A CAT can be administered on even the least expensive personal computers now available. A test has the potential to be different for each examinee depending on how they perform on the test items. For example, if an examinee displays a consistent response pattern based on his or her proficiency level, he or she may take a shorter CAT. Whereas, if an examinee displays an aberrant or irregular response pattern (e.g., answering a difficult item correctly and an easy item incorrectly) based on their proficiency level, he or she may take a longer test because it will be more difficult to estimate his or her actual proficiency level. Each item on a CAT typically has a known standardized difficulty level and discrimination index.

Conventional paper-and-pencil tests employ many items that generate little information about the examinee, especially for examinees at the extreme ends of the measured construct. For example, proficient students are typically asked too many easy, time-wasting questions. An even worse scenario may involve low-ability students being asked too many difficult questions that may detrimentally affect their self-confidence. An advantage of CAT involves the administration of only a sample of the total items in the test bank item pool to any one examinee; this reduces the number of items that need to be administered by as much as 50 percent. On the basis of previous response patterns, items that have a high probability of being answered correctly (if it is a proficiency test) are not presented, thus providing economy in terms of testing time and total number of items presented (Embretson, 1996).

Another advantage of the CAT is the utilization of an item bank, which is a set of test items stored in computer memory and retrieved on demand when a test is prepared. Each item is stored and uniquely classified by several dimensions, such as (a) item type, (b) content measured, (c) difficulty level, and (d) date of last use. According to Cohen and Swerdlik (2002), another major advantage of the CAT is the capability for item branching—the ability of the computer to tailor the content and presentation order of the items on the basis of examinee responses to previous items. For example, a computer may be programmed to not present an item related to the next difficulty level until two consecutive items of the previous difficulty level are answered correctly. The computer can also be programmed to terminate an exam or a subcategory of an exam at specified levels.

In general, the potential advantages of CAT are attributed to the objectivity, accuracy, and efficiency that computers and software bring to various aspects of testing and assessment. According to Cohen and Swerdlik (2002), CAT may include these potential disadvantages: (a) A CAT may be an intimidating experience for an examinee; (b) test-taking strategies that have worked for examinees in the past, such as previewing, reviewing, and skipping around the test material to answer easier questions first are not possible; and (c) examinees are deprived of the option to purposefully omit items because they must enter a response before proceeding to the next item. Tailored testing appears to be moving out of its experimental phase and becoming more accessible to a larger measurement audience because computer technology is cheaper to purchase and because item response theory applications are becoming easier to use.

Testing and Assessment Litigation

Another important issue for measurement professionals in the 21st century pertains to testing litigation. Cizek (1993) explained that competency tests, licensure examinations, and personnel evaluations will undoubtedly continue to be challenged in courts when opportunities for advancement are denied or bias may be present. Cizek believes that the process of setting standards will probably receive the most scrutiny, given its arbitrary nature and debatable empirical grounding in the field of psychometrics.

A final legal or ethical consideration involves the role of test publishers in assessment. Rudman (1987) believed that test publishers would continue to be plagued by the improper use of their tests. For example, he indicated that once a test company sells a test to a client, it has limited or no control over the way the test is administered or interpreted. Measurement professionals should be conscientious regarding the improper administration or interpretation of a test or assessment that can deleteriously affect examinees.

SUMMARY

During the last century, educational and psychological tests and assessments have demonstrated tremendous utility for addressing a wide range of applied problems. They have proved to be excellent tools for facilitating increased understanding through theory development and research. However, testing and measurement professionals should remain cautious because there is always the potential for tests to be used in harmful, inappropriate, or inaccurate ways. Consequently, these professionals must be familiar with, and guide their research practices in accordance with, the American Psychological Association's (APA) ethical practices as well as the American Educational Research Association, APA, and National Council on Measurement in Education testing standards (1999). Further, they also must be intimately familiar with the possible negative effects of various test uses and with procedures in which those deleterious effects can be minimized. If tests and assessments are used cautiously, knowledgeably, thoughtfully, and ethically, their potential for great benefits and wide practical utility can be fully realized.

REFERENCES AND FURTHER READINGS

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47–63). New York: Holt, Rinehart & Winston.

Cizek, G. J. (1993). Some thoughts on educational testing: Measurement policy issues into the next millennium. *Educational Measurement: Issues and Practices, 12*, 10–16.

Code of fair testing practices in education. (1988). Washington, DC: Joint Committee on Testing Practices.

Cohen, R. J., & Swerdlik, M. E. (2002). *Psychological testing and assessment*. Boston: McGraw-Hill.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: Murray.

Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement, 11*, 59–79.

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*, 341–349.

Engelhard, G., Jr. (1997). Introduction. *Educational Measurement: Issues and Practice, 12,* 5–7.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). New York: Macmillan.

Garb, H. N. (2000a). Introduction to the special section on the use of computers for making judgments and decisions. *Psychological Assessment, 12*, 3–5.

Garb, H. N. (2000b). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that! *Psychological Assessment, 12*, 31–39.

Glaser, R. (1994). Criterion-referenced tests: Origins. *Educational Measurement: Issues and Practice, 12*, 9–11.

Gronlund, N. E. (2003). *Assessment of student achievement* (7th ed.). Boston: Allyn & Bacon.

Hambleton, R. K., Swaminathan, H., & Rogers, J. H. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hofer, P. J., & Green, B. F. (1985). The challenge of competence and creativity in computerized psychological testing. *Journal of Consulting and Clinical Psychology, 53*, 826–838.

Honaker, L. M., & Fowler, R. D. (1990). Computer-assisted psychological assessment. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (2nd ed., pp. 521–546). New York: Pergamon.

Jacobs, L. C., & Chase, C. I. (1992). *Developing and using tests effectively: A guide for faculty*. San Francisco: Jossey-Bass.

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues* (5th ed.). Belmont, CA: Wadsworth & Thompson Learning.

Linn, R. L. (1989). *Educational measurement* (3rd ed.) New York: Macmillan.

Lyman, H. B. (1978). *Test scores and what they mean* (3rd ed.) Englewood Cliffs, NJ: Prentice Hall.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Ory, J. C., & Ryan, K. E. (1993). *Tips for improving testing and grading* (4th ed.). Newbury Park, CA: Sage.

Osterlind, S. J. (2005). *Modern measurement: Theory, principles and applications of mental appraisal*. Upper Saddle River, NJ: Prentice Hall.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20*, 12–20.

Popham, J. W. (1993). Educational testing in America: What's right, what's wrong? *Educational Measurement: Issues and Practice, 12*, 11–14.

Rudman, H. C. (1987). The future of testing is now. *Educational Measurement: Issues and Practice, 6*, 5–11.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*, 8–11.

Sturges, J. W. (1998). Practical use of technology in professional practice. *Professional Psychology: Research and Practice, 29*, 183–188.

Thorndike, E. L. (1910). The contributions of psychology to education. *The Journal of Educational Psychology, 1*, 5–12.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Traub, R. E. (1997). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice, 12*, 8–13.

Walsh, B. W., & Betz, N. E. (2001). *Tests and assessment* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Ward, A. W., Stoker, H. W., & Murray-Ward, M. (1996). *Educational measurement: Origins, theories, and explications: Vol. I. Basic concepts and theories*. New York: University Press of America.

Weiss, D. J., & Vale, C. D. (1987). Computerized adaptive testing for measuring abilities and other psychological variables. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 325–343). New York: Basic.

Wilbrink, B. (1997). Assessment in historical perspective. *Studies in Educational Evaluation, 23*, 31–48.

42

Personality Development

Philip Lewis

Auburn University

The topic of personality development lies at the interface of two larger and more familiar topics: personality theory and developmental psychology. However, as the present summary will demonstrate, most explorations of personality development focus on broad issues of identity and not on the narrower personality traits and dispositions that dominated the field of personality theory during the end of the 20th century. For all intents and purposes, the topic of personality development could as easily be termed "person development." In ways that will be made clear in this chapter, most people undergo significant psychological change as they age. With respect to their psychological characteristics, infants become children, who become adolescents, who then become adults. These progressive age-related changes are much more than changes in what or how much the person knows. People "grow up" as they age, and an important part of growing up entails progressive shifts in *how* they know (in contrast to *what* they know) and, in particular, how fully they understand themselves and others. Interestingly, not everyone "grows up" to the same extent or at the same rate. We all know chronological adults who are, nonetheless, psychological adolescents. Indeed, knowing where someone is on the path of increasing psychological maturity may be the most important thing we can know about a person. Developmental differences in personal maturity are differences that make a difference. This process of personality development, of growing up, is the focus of the present chapter.

Much of modern personality research has focused on personality traits and attendant dispositions to act in certain ways in social situations. This chapter provides little focus on the development of these sorts of personality traits. It's not that changes in personality traits don't occur. Operating out of the Five Factor Model of personality, Costa and McCrae (1994) concluded that modest changes in individuals' positions along five basic personality dimensions do occur but are "essentially finished by age 30" (p. 148). In contrast, increasing levels of psychological maturity, ego identity, perspective, and integrative complexity are pointedly *not* typically complete by the age of 30. This chapter focuses primarily on the development of these more basic qualities, the structural qualities of a person's identity. The nature of these qualities and the challenges of studying them can be illustrated by a brief example.

Having recently returned to school from a trip home during his first year at college, "Jason" (not his real name) is reflecting upon some recent changes that he has noticed in his relationship with his parents:

> . . . before when I was actually living at home I never wanted to spend time with my parents. I never wanted to get to know how they were doing, if they were feeling okay, how they were doing emotionally. And now I'll go home and I find myself wanting to go out to eat with them, so I can just talk with them one-on-one and get a feel of how they're doing.… And my mom, she just started a new job, and I honestly care about how she's doing. Like before, it wasn't really about

> how much I missed them. Before, I would miss them because they're my parents, and you love your parents. But now it's more of a friendship thing than just the love that comes with being the child of a given parent. Now it's more about the friendship and the bonding that I can do with them.

On the basis of these remarks, and others that were part of an assessment interview that lasted for about an hour, most developmental psychologists would conclude that Jason was beginning a period of significant personality development. Not incidentally, Jason's own excitement about these changes was palpable. Talking about why he felt good about having made a date to have lunch with his mother, he says:

> …it was good because of the fact that I wanted to do it. It was me setting it up, not her asking me to go to lunch with her.…Because before she'd beg me to go and do something with her, and I never wanted to do it. And now, I was just excited that it was me—I had the feeling to want to do it.

There are many ways that one could describe the changes Jason was experiencing. Not all of these changes can be considered "development" from a psychological perspective. Thus, it is instructive to identify the features of Jason's remarks that would or would not lead a developmental psychologist to conclude that Jason's experiences reflected personality development. Perhaps Jason has merely acquired a new social skill, asking people to go out to lunch, and now he's using that skill when he goes home to visit his parents. If so, his personality hasn't changed, only his behavior has. Or perhaps it's even simpler than that. Maybe the change Jason describes is due to a change in his circumstances, not his personality. After all, now that he's away at college he doesn't see his parents very often. So when he's home he makes an effort to spend time with his parents so he can maintain the level of contact with them to which he had grown accustomed as a high school student. Arguably, neither of these simpler explanations captures the source of Jason's excitement. In the remainder of this chapter you will learn both what this sort of personality development is all about and how some of the major theorists from the second half of the 20th century, from Erik Erikson to Robert Kegan, have shed light on the sort of changes we see happening in our young college student.

HOW DO PSYCHOLOGISTS DEFINE PERSONALITY AND PERSONALITY DEVELOPMENT?

There is no consensus among modern psychologists about how to define personality. Lay persons and some trait theorists tend to use the term *personality* to describe a set of social skills and personal attributes that affect how someone is experienced by others—as in "Jennifer has a sparkling personality" or "Philip has zero personality." Psychologists with an interest in personality development, as distinct from the development of social and interpersonal dispositions, generally favor definitions of personality development that reflect increasing psychological maturity. So, for example, while 21-year-old Jennifer may have a "sparkling personality," shouldn't we expect her sparkling personality to look quite different when she reaches middle age? If her "personality" doesn't change during that 20- to 25-year period, wouldn't that be a source of some concern? Jason may have many of the same personality characteristics he had when he was a junior high school student (he seems to be rather friendly and upbeat), but arguably something is happening that most of us would see as evidence that he is growing up. He is moving from seeing his parents as people who do things to you and for you to seeing his parents as people in their own right. Some young adults never make this transition.

WHAT IS IT ABOUT PEOPLE THAT CHANGES AS THEY GROW UP?

In recent years there has been a renewed interest in personality traits. Stimulated by the Five Factor Model of personality, psychologists have been impressed by the degree to which the trait terms people use to describe one another (e.g., friendly, extroverted, sensitive), as well as actual self-descriptions and descriptions of others' personality traits, can be summarized using five relatively independent personality dimensions. These five dimensions (often labeled Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience) appear to work well in describing individuals at different ages, and across many different cultures. However, these traits do not capture many important features of personality, and are not the features of personality that are the focus of most psychologists who study personality development. Indeed, these broad personality traits, once they become a consistent feature of someone's personality—usually in middle childhood, tend to be remarkably stable across the remainder of one's life.

What is not adequately captured by merely describing someone's personality dispositions or traits are the ways in which people *organize* their behavior, cognitions, and feelings in the process of trying to make sense of their personal, interpersonal, and cultural experiences. More than dispositions to act in certain ways, personality includes stable features of individuals' motivations, assumptions, and predictions about social reality, and strategies for adapting to the demands of the social world. In this sense, personality is much more than a set of dispositions to behave in certain ways. Personality organizes and operates on experience. It is the part of each of us that selectively attends to certain experiences and not others, attaches particular meaning to those experiences, seeks to participate in certain aspects of the social environment and not other aspects, and actively shapes and manipulates the circumstances of one's social and cultural life. These more complex aspects of personality develop and change over time.

As suggested in the interview with Jason, not all change is development. The convention that will be adopted in the present chapter is that personality development refers to certain kinds of *durable internal* changes in the person. Jason's *external* behavior toward his mother has surely changed; he now asks her to go to lunch. However, even if he were to stop asking, it appears that Jason has a different internal view or understanding of who his mother is, a person in her own right, someone who has interesting qualities apart from being the mother who takes care of him and provides him with things that he needs or desires. Jason's mother hasn't changed. She has always been more than merely Jason's mother. What has changed is Jason's capacity to construct this more encompassing view of his mother. It is a view that is not likely to be lost. Once one can see more of something that was there all along, it's pretty hard to stop seeing it. Jason has undergone personality development. As a result of his new capacity to see another person more fully, we can expect to see additional evidence in his behavior of this enduring internal developmental change.

One final feature of personality development is that it entails capacities that enable the person to adapt to or cope with environmental demands and environmental complexity. Jason's newfound capacity to see beyond his own needs in constructing an understanding of what other human beings are all about will help him be more successful in relationships beyond that with his mother. For example, given the opportunity to exercise his new capacity, Jason will be better able to appreciate the personal qualities of a future spouse, as well as what it means to become a "professional" rather than merely someone who does a particular job. Jason will be more successful or "better adapted" in both endeavors, marriage and vocation, because he can appreciate people in ways that are unconnected to his need for them.

KEY METHODOLOGICAL ISSUES WHEN STUDYING PERSONALITY DEVELOPMENT

It may seem too obvious to mention, but personality development is about psychological change. As a result, one of the challenges in investigating personality development is deciding how best to study change. Assume for a moment that you have become interested in the sort of change that our college student, Jason, appears to be experiencing. You begin wondering if most 19-year-olds have a similar experience. They start becoming interested in their parents' lives and feelings in a way that is no longer connected to what their parents can provide or withhold. You have several options about how you might proceed. If you want to get some quick answers to your question, you could proceed by assessing the parental views of three groups of individuals, perhaps 15-year-olds, 20-year-olds, and 25-year-olds. If you are indeed measuring a variable that is a feature of adolescent or young adult personality development, then you would expect the older children's views of their parents to become less about their own needs and more about their parents' personal experiences. This investigative approach is termed "cross-sectional" and has the advantage of providing you with answers relatively quickly.

The cross-sectional approach has drawbacks, though. One drawback is that you can never be sure that the age differences you find are in fact a function of development. For example, what if it is the case that back when your 25-year-olds were in their late teens, TV was filled with highly popular sitcoms that focused on family life. Subsequently, the popularity of such TV programs faded and family-oriented programs were replaced by a variety of programs focused on competition and individual achievement, such as "Survivor" and "American Idol." As a result, your 25-year-olds spent many hours watching programs about parents and children, while your 20-year-olds spent many hours watching programs about advancing one's own goals at the expense of others. So even though your 25-year-olds were much more attuned to their parents' experiences than were your 20-year-olds, the difference could be due to the two groups' having been exposed to different types of popular TV programs.

To sort this problem out, you would need to turn to the other major experimental method used to study development, the "longitudinal" approach. In this approach you would assess your participants at two or more points in time. You might, for example, want to reassess your 20-year-olds when they were 25. Do they show evidence of becoming more attuned to their parents' inner lives than they did when you assessed them five years earlier? If so, then you have more confidence that you have detected personality development.

The longitudinal approach also has its drawbacks. Most obvious is that it requires more time and effort than the cross-sectional approach. In our example, you had to wait five years before you were able to find out if there had been any change with respect to the parental views of your longitudinal study participants. You would also have to make sure you could locate all of your participants five years after first assessing their views and then convince them to take the time to be reassessed. Even if they agreed, it is possible that their responses might have changed merely because they were familiar with your study and were being asked the same questions a second time. Notice that you didn't have to deal with these potential complications when using the cross-sectional approach.

An even more serious complication of the longitudinal approach may result if you have a significant amount of "attrition" from your original sample—if you are unable to reassess a significant number of your original participants. Let's say that, for some reason, it was much harder to recontact participants who were highly focused on themselves and their own individual interests. Perhaps these self-interested participants were more likely to move out of the area and away from their families or were still around but were much harder to convince to participate in your study a second time. A loss of a particular type of participant would be termed "differential attrition" and would make it very difficult to interpret your longitudinal results. If you found that at age 25 your remaining participants were more attuned to their parents' inner experiences than your participants had been at age 20, it could be due to the fact that you had lost from your sample of 25-year-olds those individuals who would have still been focused upon themselves. So one thing you need to be alert to when reading the results of longitudinal studies is whether there is attrition, and if there is, whether the researcher presented information that suggests the attrition was random as opposed to "differential." One common way in which researchers try to solve the problem of differential attrition is by presenting only the longitudinal results for individuals who participated throughout the entire study. However, even in this case one needs to be aware that the results reported may not apply to individuals who had characteristics that caused them to drop out of the study.

There is a final methodological issue relevant to investigating personality development using a longitudinal approach, and that concerns whether one chooses to look at individual or group data. If you look only at group effects, you lose the ability to learn something about the process by which individuals change, or as is often the case, fail to change. As some developmental researchers have noted, it may be the case that rather than taking place through a fixed set of progressive stages, some personality development may be better described as taking place through a number of different "trajectories" or "pathways" (Bosma & Kunnen, 2001).

We turn now to brief summaries of some of the more prominent theories of personality development that incorporate the sorts of changes that are taking place in Jason. As you will see, there is considerable variation in how different theorists have conceptualized these enduring, internal, and adaptive changes in the person.

ERIK ERIKSON AND IDENTITY DEVELOPMENT

Child psychoanalyst Erik Erikson put forth one of the oldest and best-known theories of personality development. Writing in the 1950s and 1960s, Erikson (1968) provided a stage theory of identity formation and consolidation in which he described eight universal identity crises that occur in a fixed sequence across the human life span. Each identity crisis is precipitated by changes in cultural expectations that correspond to expected psychological and biological changes in the person. For example, in Erikson's Stage 5, which typically occurs somewhere between the ages of 13 and 20, depending upon various cultural and biological factors, most cultures expect children to begin to act like adults. In Western cultures we refer to individuals in this age range as adolescents, and we expect them to begin acting like productive members of society by getting an education or finding a vocation, dating or marrying, and showing a new level of concern about how their current decisions and actions might impact future life success.

Erikson believed that viable societies begin making these demands on adolescents just at that point in the life span when biological and psychological changes are taking place that provide adolescents with a good chance of successfully meeting these demands. After all, Erikson noted, societies have a vital interest in producing new members who can ensure their continued success. Among the personal changes that help most adolescents cope with these societal demands are a new capacity for self-reflection and an ability to internalize others' views. According to Erikson, the predictable identity crisis of adolescence results from the confluence of these internal and external changes. In response to these emerging internal capacities and growing external pressures, adolescents must figure out who they are to become and whether that new identity will be recognized and supported by others. Erikson argued that whether adolescents will be successful in forging this new, more adultlike identity depends both on the quality of the societal support provided and on the strengths the adolescent can bring to bear on the current crisis. These strengths arise from the successful resolution of preceding identity crises. Three succeeding identity crises, focused in turn on achieving personal and vocational commitment, selflessness, and life integration, will similarly result from predictable internal changes and external pressures. Erikson believed that each succeeding identity crisis offered the opportunity for one to address earlier identity issues if those identity issues were not successfully resolved when they were first encountered.

Erikson's developmental theory both is well known and has stimulated considerable related research. The most well-known derivative of Erikson's views is the theorizing and related research program initiated by James Marcia (1993). Drawing upon Erikson's formulation of the Stage 5 adolescent identity crisis, Marcia described an "identity status" model in which he classified adolescents as occupying one of four statuses depending on where they are with respect to the resolution of their adolescent identity crisis. Marcia's "diffused" status obtains where no exploration has taken place and no identity has been achieved. His "foreclosed" status is characterized by a superficial commitment to an adolescent identity without going through the process of personally struggling with identity issues. These adolescents may, for example, unquestioningly adopt their parents' values and expectations. The "moratorium" status exists when active exploration of various identities is underway but the adolescent has not yet committed to a single identity. The most evolved status is "achievement," where the adolescent has committed to a particular identity after an extended process of personal exploration. Research has focused on both confirming the existence of the four identity statuses as well as exploring the process by which some adolescents move through the less satisfactory statuses to "achievement" while others appear to get stuck or cycle repeatedly between diffusion and moratorium (for a review, see Bosma & Kunnen, 2001). Little or no research has been focused upon the important question of what facilitates transitions from one identity status to another.

JANE LOEVINGER'S THEORY OF EGO DEVELOPMENT

In the 1960s and early 1970s, Jane Loevinger took a somewhat different approach to describing and investigating the lifelong process that she called "ego development." Though nominally related to Freudian psychoanalytic theory, Loevinger's concept of ego development is in fact quite different from most psychoanalytic views because it refers to the changes that occur in how a person constructs meaning. She distinguished ego development from both intellectual development and level of psychological adjustment. Her view is that as we progress from early childhood through adulthood we have the potential to construct a progressively more complex and encompassing understanding of who we are in relation to others. In her view, young children derive meaning from the satisfaction of their immediate needs and desires, and others are viewed in black-or-white terms with respect to whether they give the child what he or she desires at the moment. Psychologically mature adults, in contrast, make sense of themselves in terms of enduring internal qualities and allegiances that transcend immediate needs or desires, and mature adults are also able to recognize others as having similar inner qualities that exist independent of one's relationship to these individuals. Between these two poles of meaning construction is a great deal of "ego development."

As is true of most comprehensive theories of personality development, Loevinger concluded that rather than taking place gradually, ego development takes place in spurts or stages, where stages of ego development represent relatively stable periods of meaning making. At each successive stage, the individual has developed a new, more complex way of constructing meaning. Experiences of the self (one's thoughts, feelings, and behaviors) and interactions with others are made sense of using the current stage's framework for making sense of those experiences. Experiences and interactions that don't fit the current sensemaking framework are either ignored or distorted to fit the current framework.

It is this so-called "selective gating" that confers the developmental stability that results in stages. Only when experiences of the self and others that cannot be adequately understood using the frames of one's current ego development stage meet some threshold, and can no longer be ignored or distorted, does the person enter a transition phase on the way to the next ego development stage, a stage that will allow the person to make better sense of the discrepant experiences. In her most recent summary of the stages (Hy & Loevinger, 1996), Loevinger described eight stages, E2 through E9, and acknowledged, but did not describe, an E1 infancy stage. Loevinger arrived at her stages without using an a priori theory about how many stages there would be or what the different stages would look like. Instead, she used a "bootstrapping" method, whereby experts in personality development took the responses of a number of participants ranging in age from early adolescence through adulthood and created groupings of responses that seemed both similar to one another and significantly different from other groups of responses. When there was general agreement that participants' responses had been placed into distinct and conceptually meaningful groups, Loevinger used both participants' ages and judgments about the progressive sophistication or complexity of the response groupings to order the eight developmental stages. In the initial version of her stage theory (Loevinger & Wessler, 1970), the major stages were given a number and a capital "I" (I1 through I6) and two transition stages were designated with two adjacent stage numbers (I-3/4 and I-4/5). If you read the early literature on ego development, you may encounter these "I" designations rather than the "E" stages of the current theory.

We can use Jason, the college freshman whose interview was excerpted at the beginning of this chapter, to get a better feel for what Loevinger's ego development stages describe. We will need to examine the content and structure of Jason's remarks to make a guess about his current stage of ego development. It is worth noting that examining Jason's statements was not necessary when we were considering Erikson's theory of identity development. Merely knowing that Jason was 19 years old and a member of a Western culture, we would expect Jason to be finishing up Erikson's Stage 5 (Identity vs. Role Confusion) and entering Stage 6 (Intimacy vs. Isolation) because biological maturation and societal expectations are expected to precipitate each of Erikson's identity crises. Loevinger's theory has a different view of the impetus for stage progression: One must encounter a sufficient level of discrepant experiences to precipitate development to the next stage. If Jason has been underchallenged, then he could be at a developmentally lower or earlier stage than another 19-year-old who has had to deal with a greater number of life experiences that he or she could not easily assimilate to the sensemaking framework of a lower stage. We will consider three of Loevinger's stages that could describe Jason: E4, E5, and E6.

Individuals at Loevinger's E4 ("Conformist") stage are solidly identified with their dominant reference group. They have a simplistic, conventional view of rules, and they believe there is a right and a wrong way of doing things. Expectations that one should follow rules and meet social expectations are thought to apply to almost everyone. Social approval for following the rules and acting appropriately is highly valued, and social disapproval for not doing so is avoided. Individuals at this stage are said to be preoccupied with surface characteristics such as one's appearance, possession of "cool stuff," and membership in the right groups and cliques. Inner experiences are described in rather simplistic terms (e.g., happy, sad, mad, excited, bored) and others are viewed in equally simplistic terms. There is little in Loevinger's description of E4 that seems to capture Jason's concerns. Jason seems uninterested in having his mother's approval and there is no sense that he is asking her to lunch and taking an interest in her experiences because that is the "right" thing for a son to do. Jason is probably not making sense of his new experience of himself in relationship to his mother using the framework of Loevinger's E4.

Individuals at Loevinger's E5 ("Self-Aware") stage have begun to realize that people have characteristics that exist outside of surface characteristics and their conformity (or lack of conformity) to the expectations of society or one's group or clique. This realization leads to a new interest in internal experiences as something that can vary widely even among members of the same group. Often there is a growing awareness of the differences between the self and members of groups with which one was closely identified at the previous stage. As a result, there is a growing sense that there are legitimate exceptions to the rules and to the value of social conformity. Loevinger's E5 stage seems much closer to the experiences Jason is describing. He is interested in getting to know more about his mother's experiences in her new job quite independent of her role as his mother. Jason is also intensely interested in his own inner motivations and his changing experience of his mother, but before we conclude that Jason is functioning at Loevinger's Stage E5, let's consider her description of the next stage.

Individuals at Loevinger's E6 ("Conscientious") stage evidence movement away from a belief that there are certain rules and social expectations to which everyone should adhere and toward the adoption of self-evaluated standards. Acceptable behaviors are not just a function of what others think is appropriate but are increasingly judged acceptable or unacceptable using one's own personal feelings. Because what is acceptable has moved inside the person at E6, it now becomes important, in judging another's actions, to know what the person's motives for those actions were. Merely having broken a rule is not nearly as important as knowing how one's rule-breaking made another person feel. Individuals functioning at E6 are self-reflective and are likely to describe their inner experiences in great detail. Because these individuals make decisions based on internal standards, they recognize multiple possibilities, and often describe a sense of having to make choices based on identifiable personal reasons. There is an orientation toward long-term goals and gradual self-improvement. There would appear to be little in Jason's remarks that reflect the core issues and concerns of this stage. Although it is possible that Jason might have expressed different concerns, were he not talking about his new interest in spending time with his mother, Loevinger's E5 ("Self-Aware") stage seems to best capture how he is struggling to make sense of his experiences in a new, more complete manner.

At this point, you may be wondering how Loevinger assesses an individual's stage level. In much of personality research, particularly with respect to trait and dispositional approaches such as the so-called "Big-Five" model described at the start of this chapter, researchers rely on self-report questionnaires. From Loevinger's perspective, the main problem with this approach is that it is the questionnaire's author who creates the response categories, not the person being assessed. Because Loevinger is seeking to determine how an individual imposes meaning upon the experiences he or she is encountering, she uses a free response or "projective" approach to assessment. Her primary assessment tool is a 36-item sentence completion form on which participants write responses that complete sentences such as "The thing I like about myself is…" and "Rules are…" (Hy & Loevinger, 1996, p. 28). Loevinger provides an extensive scoring manual where one determines a person's developmental stage by matching each respondent's sentences with examples determined to be typical of individuals at each of her ego development stages. Such an approach to assessment is not without problems, such as the need to establish interrater agreement, but its strength is that the sentences are created by the respondent and not the assessor, which presumably increases the likelihood that one is getting a reasonable snapshot of how the individual being assessed imposes meaning on his or her experiences. Because Loevinger has provided an accessible, if somewhat demanding, assessment tool for assessing a person's ego development level, her approach has generated a reasonable amount of empirical study (Westenberg, Blasi, & Cohen, 1998).

ROBERT KEGAN'S THEORY OF THE DEVELOPMENT OF THE SELF

One of the weaknesses of Loevinger's theory of ego development is its lack of an integrative theory of ego development. Despite the fact that Loevinger described the ego as a "master trait" that provides overall coherence to one's experience of self and others, she never provided a sound theoretical basis for this view and what it is in the person that generates greater and greater ego maturity. This shortcoming is evident in her admission that her eight stages may include transitions as well as stages, but that it is difficult to distinguish between the two.

Beginning in the 1980s, Harvard psychologist Robert Kegan outlined a theory of the development of the self that appears to have the potential to overcome these theoretical limitations. Drawing on an implicit feature of Piaget's developmental theory, Kegan described six stages in the development of the self that are based on progressive transformations of the developing person's perspective-taking capacities. Following an infancy stage, Stage 0, in which the infant lacks the capacity to distinguish between self and other and there is no coherent perspective taking, the toddler entering Stage 1 makes sense of her or his experience using a single, coherent perspective. Understanding that others have a separate and potentially different perspective from one's own does not emerge until middle childhood, as Kegan's Stage 2.

By late adolescence an even more complicated perspective-taking capacity, the capacity to hold two or more perspectives together simultaneously, becomes the fundamental basis of Kegan's Stage 3. As with each successive gain in perspective-taking capacity, this new ability to experience multiple perspectives at the same time again transforms the developing person's sense of the self and others. In Kegan's view, the capacity for simultaneous perspective-taking permits the Stage 3 individual to apprehend complexities that were present but could not be appreciated using the simpler perspective-taking capacities of earlier stages. Stages 4 and 5 are characterized by the emergence of even more complicated perspective-taking capacities. The way in which Kegan's theory provides a better grounded basis for describing developmental stages than does Loevinger's bootstrapped approach can be seen by using Kegan's theory to again analyze the remarks of our college student, Jason.

Research using Kegan's theory suggests that college freshmen are likely to be in Kegan's Stage 2, Stage 3, or in the transition between those two stages. Let's start with Stage 3. Kegan asserts that the capacity to take two distinct perspectives and hold them together at the same time, which is the new perspective-taking capacity at Stage 3, creates certain new personal and interpersonal experiences.

Taking the interpersonal first, if Jason were functioning at Stage 3 we would expect him to show evidence that he was beginning to make his mother's experience of him (one perspective) a part of his experience of himself (another perspective, taken simultaneously with the first). He might say something like "When she used to beg me to go and do something with her and I'd refuse, it would hurt her feelings. Thinking about that now makes me feel bad." He would be taking her feeling bad and making it a part of his own feelings about himself. However, if you look carefully at what Jason says, he doesn't, in fact, feel bad as a function of his mother's feelings about him. Instead, his mother's desires—"before she'd beg me to go and do something with her"—are viewed by Jason as an alternative to his own desires—"and I never wanted to do it." He appears not able to incorporate how his refusal might make his mother feel as a part of his own experience of himself. Thus, at the interpersonal level, Jason appears to still be functioning at Kegan's Stage 2. He clearly perceives there to be two different perspectives or desires (his mother's and his own), so he is not functioning at Stage 1, but the two points of view remain entirely separate in his experience of them.

Now let's consider the intrapsychic aspects of Jason's remarks. At Stage 3, rather than merely being aware of one's own feelings and thoughts, which is a Stage 2 capacity, one begins having feelings about one's feelings and thoughts about one's thoughts. In other words, Kegan suggests that the Stage 3 capacity to hold two thoughts or two feelings together at the same time (another expression of simultaneous perspective taking) creates, for the first time, self-reflectiveness, or what clinical psychologists sometimes refer to as "psychological mindedness." In this respect, Jason appears to be moving a bit beyond Stage 2. Indeed, what has struck Jason as new and different is the way he is experiencing his experience of his mother. As he puts it, "…it was good because of the fact that I wanted to do it…and now, I was just excited that it was me—I had the feeling to want to do it." In short, Jason is telling us that he is excited about how he is feeling. He's having feelings (excited) about his feeling (wanting to be with his mother), which is evidence that at the internal or "intrapsychic" level, Jason is becoming more Stage 3. Indeed the entire interview from which the excerpt is taken places Jason in the Stage 2 to Stage 3 transition.

What distinguishes Kegan's theory from Loevinger's is his use of a single theoretical basis—perspective-taking complexity—for describing each of his developmental stages. Lacking an overall theoretical basis for identifying her stages, Loevinger is left having to describe a collection of behaviors, thoughts, and feelings that characterize each stage without having any clear sense of what it is that produced each stage's particular features. In contrast, Kegan is able to use his underlying theoretical framework to provide an explanation for why many of the features Loevinger locates together in each of her stages can be expected to occur together. Kegan's framework can also be used to identify transition phases that, from his perspective, are distinct stages. In this way, one can reduce Loevinger's 10 stages to 6 stages and 4 transitions that correspond closely to Kegan's six stages.

Kegan's theory is not without its limitations, however. Foremost is the difficulty of his method for assessing individuals' developmental levels. Kegan agrees with Loevinger that because one is assessing how individuals make sense of their experience, questionnaires will not provide valid assessments. Instead, like Loevinger, Kegan favors a free response approach, where the person being assessed generates his or her own responses to open-ended stimulus materials. However, whereas Loevinger relies on sentence stems for her assessments, Kegan uses open-ended, face-to-face interviews, and asks respondents to describe recent experiences that evoked strong emotional reactions. The interviewer then follows up the respondent's descriptions of those emotionally laden situations with a series of "Why did you feel that way?" questions designed to elicit the person's perspective-taking capacities. It is a process that requires extensive training, and establishing adequate interrater agreement is sometimes hard to achieve. Despite its challenges, Kegan's theory is slowly attracting the interest of other scientists, most notably organizational psychologists like Wifred Drath at the Center for Creative Leadership and William Torbert, former dean of the business school at Boston College. Both have written books that draw upon Kegan's developmental theory in trying to make sense of leadership and organizational behavior.

ADDITIONAL APPROACHES TO THE STUDY OF PERSONALITY DEVELOPMENT

There are many approaches to the study of personality development in addition to Erikson's, Loevinger's, and Kegan's. But unlike these three approaches, none are as comprehensive, and some may not even qualify as developmental approaches with respect to the way I have defined personality development in this chapter. Some approaches have a more restricted age range; some approaches focus on a particular gender; other approaches focus on only part of the life span; and still other approaches focus on a single, narrow aspect of personality, such as moral decision making or conceptual development. We will begin this brief overview with some of the approaches that have the longest history, move to approaches with a narrow focus, and finish with two of the more promising approaches that have emerged in the past few years.

Historically Important Theories

Daniel Levinson (1968) and his colleagues at Yale University utilized open-ended interviews to identify patterns of personality change from late adolescence through middle adulthood. His primary focus is on what he called the changing "life structures" that a person constructs to organize his or her approach to the tasks typically encountered during each of adult life's changing "seasons." Levinson provided a rich description of the many struggles that occur as one moves from the challenges of establishing financial and emotional independence in early adulthood to building a career and family during middle adulthood and later reevaluating and reassessing previously held dreams and aspirations. In many respects, Levinson's work is less theoretical than descriptive, and it is not always clear whether he is describing changing patterns of adaptation or changes in personality. Like our next theorist, Levinson was influenced by the work of Erik Erikson.

George Vaillant (2002) is best known for his careful analysis of several in-depth longitudinal studies of Harvard University students, which followed samples of graduates for as long as 45 years. In an approach similar to Erikson's, Vaillant identified a series of developmental tasks that people typically encounter as they move through adulthood. As a psychiatrist, Levinson was interested in what distinguished psychologically well-adjusted graduates from those who were adjusting less well to adult life. A developmentally significant feature of this focus on adjustment was Vaillant's attempt to describe six increasingly mature mechanisms for dealing with the anxieties and challenges of adulthood. At the highest level was what he termed "altruism," where one attempts to reduce anxiety and uncertainty by devoting oneself to helping others, which contrasts with a lower level strategy, "autistic fantasy," where one deals with anxieties by daydreaming about favorable outcomes rather than taking action to confront the sources of one's anxieties. Successful development entails meeting the demands of adulthood's predictable challenges and developing the inner resources to be able to do so with a manageable level of personal distress. Like Levinson's work, Vaillant's work is as much descriptive as it is theoretical, and is focused as much on changing patterns of adaptation as it is on changes in personality.

Women's Personality Development

Using a combination of longitudinal and cross-sectional methods, University of California professor Ravena Helson (1993) and her colleagues followed the adult life experiences and personality changes in two sets of Mills College graduates. Like Levinson's and Vaillant's mostly male participants, Helson's female college graduates went through significant periods of reevaluation and adjustment in their views of their lives and themselves. Many of these well-educated women focused considerable energy in their 30s and early 40s on the issue of forging an independent, self-authored identity, a quest that was accompanied by a decline in typical measures of femininity. The successful women described themselves as developing greater self-confidence, independence, and self-definition. They showed a concomitant increase in coping skills and the emergence of long-term commitments to cherished values. Helson pointed out that the pattern of changes and struggles she observed are, in some important respects, different from those experienced by men.

Although much more limited in its scope than Helson's work, Harvard professor Carol Gilligan (1982) provided an influential theory of what she sees as the central developmental issue of women's adult development. Gilligan suggested that, in part because of culturally based gender expectations, most adult women come to feel torn between caring for others and promoting one's own self-development. Gilligan's theorizing provides a possible theoretically based way of understanding what Helson sees as women's striving for self-definition.

Another researcher who has focused on the critical issue of the development of self-authorship in early adulthood is Marcia Baxter Magolda (1998), a professor of educational leadership at Miami of Ohio. Baxter Magolda's assumption is that our society demands that young adults begin to take personal responsibility for managing their lives and careers. Drawing upon Kegan's theory of the development of the self, Baxter Magolda used in-depth interviews to illuminate ways in which naturally occurring challenges and supports as well as intentional educational programs can facilitate developmental change in both men and women. Baxter Magolda's assertion that self-authorship is a common feature of development during the college years is at variance with the findings of the present author and his colleagues (Lewis, Forsythe, Sweeney, Bartone, & Bullis, 2005), who assessed the self-development of a sample of West Point cadets. Their longitudinal findings suggest that the key developmental issue during the college years is the establishment of a shared identity and finding oneself within groups and relationships rather than defining oneself as distinct from those connections. The difficulty of mastering Kegan's assessment method may be contributing to the different conclusions of these two research programs.

Promising New Approaches

Better known to educators than psychologists, Patricia King and Karen Kitchener (1994) formulated a stage theory describing the developmental process by which people are able to make increasingly sophisticated interpretive judgments and present well-reasoned arguments about complex issues and problems. According to King and Kitchener, this developmental process cumulates in the ability to engage in "reflective judgment" that entails the ongoing evaluation of beliefs, assumptions, and hypotheses against existing data and against alternative plausible interpretations of those data. At the earliest developmental level (Stages 1 and 2), knowledge is assumed to be absolute and concrete. Assertions are either true or they are false, and authorities, like teachers and subject matter experts, have access to the truth. At these early developmental stages most issues are assumed to have a single right answer, so there is little or no room for uncertainty. In contrast, individuals functioning at King and Kitchener's highest self-reflective stages (Stages 6 and 7) understand that knowledge is constructed, not absolute. What is "true" may depend upon the context, and the adequacy of any answer is based on what seems most reasonable according to the available evidence. Though more narrowly focused than theories of identity or the self, King and Kitchener have begun to move past mere description to focus on the critically important issue of how one promotes development.

An approach to the process of identity formation and development that has gained considerable prominence in recent years is what has come to be known as "narrative identity formation" (McAdams, 1999). Bringing together the work of many researchers, this approach focuses on the ways in which individuals employ narratives or personal stories in an attempt to construct a sense of unity and meaning for the ongoing process of their lives. Much of the focus of the narratives that have been investigated thus far concerns self-understanding and the influence of relationships, family, culture, class, and ethnicity on personal identity. This approach to studying personality development appears to be generating more empirical study than any of the other approaches to personality development included in the present chapter. Its principle weakness is that it lacks an underlying organizing theory or framework and thereby is at risk for being descriptive of the personality development process without providing an explanation for how and why that process takes place as it does.

SUMMARY

To a great extent, the topic of personality has existed largely on the periphery of mainstream psychology. Modern psychology is not much taken with the grand theories of human experience that were so popular in the first half of the 20th century. Nonetheless, the field has progressed, and with the advent of Kegan's integrative theory of the development of the self and the considerable interest generated by the narrative identity approach, there may at some point be a resurgence of interest in trying to understand better the process by which personality develops and human beings become more psychologically mature as they move through their lives. There is emerging evidence to suggest that by adulthood there is an increasing divergence between one's age and one's maturity level (Westenberg & Gjerde, 1999). Many adults just don't seem to "grow up," or at best their development slows dramatically. Because one's level of personality development appears to have a major impact on so many aspects of one's life, society has a vested interest in better understanding the process of personality development and, eventually, learning how to maximize it. For these reasons alone there is every reason to believe that the study of personality development will gain greater prominence in the future.

REFERENCES AND FURTHER READINGS

Baxter Magolda, M. B. (1998). Developing self-authorship in young adult life. *Journal of College Student Development, 39,* 143–156.

Bosma, H. A., & Kunnen, E. S. (2001). Determinants and mechanisms in ego identity development: A review and synthesis. *Developmental Review, 21,* 39–66.

Costa, P. T., & McCrae, R. R. (1994). Stability and change in personality from adolescence through adulthood. In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 139–150). Hillsdale, NJ: Lawrence Erlbaum.

Erikson, E. (1968). *Identity: Youth and crisis.* New York: W.W. Norton.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development.* Cambridge, MA: Harvard University Press.

Helson, R. (1993). Comparing longitudinal studies of adult development: Toward a paradigm of tension between stability and change. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lines through time:*

Personality and development (pp. 93–120). Washington, DC: American Psychological Association.

Hy, L. X., & Loevinger, J. (1996). *Measuring ego development* (2nd ed.). Mahwah, NJ: Erlbaum.

Kegan, R. (1982). *The evolving self: Problem and process in human development.* Cambridge, MA: Harvard University Press.

Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment.* San Francisco: Jossey-Bass.

Levinson, D. (1986). A conception of adult development. *American Psychologist, 41,* 3–13.

Lewis, P., Forsythe, G. B., Sweeney, P., Bartone, P., & Bullis, C. (2005). Identity development during the college years: Findings from the West Point longitudinal study. *Journal of College Student Development, 45,* 357–373.

Loevinger, J., & Wessler, R. (1970). *Measuring ego development* (Vol. 1). San Francisco: Jossey-Bass.

Marcia, J. E. (1993). The status of the statuses: Research review. In J. E. Marcia, A. S. Waterman, D. R. Matteson, S. L. Archer, & J. L. Orlofsky (Eds.), *Ego identity: A handbook for psychosocial research* (pp. 22–41). New York: Springer-Verlag.

McAdams, D. P. (1999). Personal narratives and the life story. In L. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 478–500). New York: Guilford Press.

Vaillant, G. E. (2002). *Aging well.* New York: Little, Brown.

Westenberg, P. M., Blasi, A., & Cohn, L. D. (Eds.). (1998). *Personality development: Theoretical, empirical, and clinical investigations of Loevinger's conception of ego development.* Mahwah, NJ: Erlbaum.

Westenberg, P. M., & Gjerde, P. F. (1999). Ego development during the transition from adolescence to young adulthood: A 9-year longitudinal study. *Journal of Research in Personality, 33,* 233–252.

43

Personality Psychology

Peter J. Giordano
Belmont University

Why does the pop star Madonna challenge religious convention in some of her songs and performances? What gave Dr. Martin Luther King, Jr., the strength of character to lead the civil rights movement despite such strong opposition and danger to his life? If we survey young people involved in high school leadership positions, can we predict who among them will be the next president of the United States? Questions like these are at the heart of the psychology of personality. Most people are implicit personality theorists (Anderson, Rosenfeld, & Cruikshank, 1994) because they consistently seek to explain why people do what they do. The theories explored in this chapter are more formal systems for trying to understand the remarkable diversity of human personality. In fact, it is this diversity, or these individual differences in behavior, that drives ongoing thinking and research in this area. The differences compel us to ask why. Plain and simple, these theoretical ideas are interesting, often controversial, and may be what first drew you to the academic study of psychology.

This chapter surveys the major theoretical perspectives on human personality. In so doing, I discuss the most important theoretical constructs within each perspective, the research methods and evidence marshaled in support of the theory, some related theoretical positions, some applications of the theory, and some comparisons across theories. For each theoretical perspective, I also briefly evaluate the contributions the viewpoint makes using criteria that are generally accepted as helpful when evaluating any theory:

1. *Comprehensiveness*: The extent to which a theory explains a broad range of personality phenomena. Sometimes this criterion is also referred to as the explanatory power of a theory. Other things being equal, the more comprehensive theory is preferred.
2. *Parsimony*: The simplicity of a theory. When comparing theories with equal explanatory power, the more parsimonious theory is preferred. This principle is sometimes called Occam's Razor, after the 14th-century English Franciscan monk who popularized the idea.
3. *Usefulness*: The degree to which a theory is helpful in the sense of having important practical applications, such as a method of behavior change (i.e., therapeutic strategies).
4. *Heuristic Function*: The degree to which a theory guides or influences future research. A theory that generates a good deal of research or stimulates lively theoretical debate serves an important heuristic function in science.
5. *Empirical Validity*: The degree to which a theory is supported by scientific research. A subtle but important aspect of a theory's empirical validity is the ability of the theory to specify the conditions under which it can be shown to be false. This principle, sometimes referred to as the criterion of refutation (or falsifiability), is a lofty and important element in a theory's empirical validation (Popper, 1959).

Few theories will be strong in every criterion. However, evaluating theories with these standards provides a good indication of the overall value and importance of the perspective.

FREUD'S LEGACY: THE PSYCHOANALYTIC TRADITION

Freud's ideas regarding the shaping of our personalities have so much filtered into our belief systems that we may think in Freudian ways without knowing it. Take the idea, for example, that our parents and home life have had a significant impact on the formation of our personalities. This position is so widely accepted today that we may not question it or may not seriously entertain the strength of other personality influences (Cohen, 1999; Hamer & Copeland, 1998; Harris, 1998). Freud is likely the chief architect of this belief, though many of us do not realize it.

Keep in mind, too, when reading this short summary that Freud developed his ideas over the course of many years of thinking and writing. Most scholars (e.g., Gay, 1988) believe that his first important work on personality was *Studies on Hysteria*, coauthored with Josef Breuer in 1895. Freud worked and wrote until his death in London in 1939. Reflecting on the time span of Freud's life reveals that his historical context contained many significant events that profoundly affected his theorizing over time. The events leading up to and surrounding the two World Wars, for example, helped shape Freud's thoughts on the destructive aspects of human nature. Others (Gay, 1988; Sulloway, 1979) have speculated that the zeitgeist of Victorian Europe as well as the interpersonal dynamics of Freud's childhood home played a role in his theorizing on the sexual forces in human personality.

Important Theoretical Constructs

Before we consider some specific ideas in Freud's approach, we need to understand the nature of constructs. Constructs are the building blocks of theories and, by definition, are complex unobservable variables (American Psychological Association [APA], 2007). For instance, anxiety may serve as a construct in a theory. Anxiety itself is a complex idea and is not directly observable; we can observe behaviors *associated with* anxiety (e.g., increased heart rate, increased perspiration, fidgety behavior, or a verbal report of "I feel anxious") but we do not directly observe anxiety itself. We infer the existence of constructs from observable behavior. We must *infer* that someone is anxious by observing their behavior and by listening to their words.

Let's now explore four overarching ideas that capture the essence of Freud's work. First, Freud assumed that human nature was characterized by primitive, lawless, biologically based drives of sexuality (Eros) and aggression (Thanatos), the so-called dual instinct aspect of his theory. Present at birth, these instincts are the foundation on which an individual personality is built.

Second, Freud emphasized unconscious determinism when understanding the complexity of personality. Determinism asserts that behavior is not random; it happens for an identifiable reason. The causal agents or determinants that drive behavior are deeply buried within the psyches of individuals, and therefore outside their awareness (i.e., the determinants of behavior are unconscious).

Third, Freud believed the human psyche was characterized by intrapsychic (within the psyche) conflict. The essential nature of the human mind, therefore, is one of dynamic forces in constant battle for ascendancy. One arena of this conflict plays out in the interplay of what Freud saw as the three structures of personality: the id, ego, and superego. The term "structure" is problematic because it implies a physical reality in the same sense that a brain structure (e.g., the hypothalamus) implies a living, physical portion of the brain. The term "agency" or the phrase "set of functions" might better capture the nature of these structures. Present at birth, the most basic of these structures is the id, which contains the dual instincts of Eros and Thanatos. Deriving from the id, the ego and superego develop. The superego corresponds to our set of moral and ethical principles that we have introjected from parents during psychosexual development, whereas the ego serves the complex task of mediating the demands of the id, superego, and reality, no easy psychological task. Thus, the psychological stage is set for the competing struggles among these various psychic agencies and between Eros and Thanatos.

Fourth, the most relevant developmental influences on personality occur in early childhood, particularly during the first five or so years of life, roughly speaking. Early relationships with parents, therefore, take on particular significance. In fact, in highlighting this issue, Freud borrowed a phrase from William Wordsworth's (1802/2002) poem "My Heart Leaps Up" and asserted "the child is father [parent] of the man [person]." Adult behavior involves playing out dramas that have their origins in this early developmental period. In terms of Freud's ideas on psychosexual development, by the age of five or six, the child would have passed through the oral, anal, and phallic stages. Lying ahead are the latency period, corresponding roughly to ages six to puberty, and the genital (or adult) stage of development. Successfully navigating these developmental stages gives rise to the development of ego and superego functions, as psychic energy or libido gets redistributed to fuel the psychological functions of ego and superego.

Related Theoretical Approaches

Because Freud's thinking was so comprehensive and because his personal conviction was so forceful, he stimulated others to revise or extend his thinking. These "spin-off" perspectives diverged from Freud's original ideas in important ways, but they still emphasized the role of the unconscious in determining behavior and the relative importance of early childhood events in shaping personality.

Carl Gustav Jung (Analytical Psychology) and Alfred Adler (Individual Psychology) were two of Freud's early adherents who split from him to promulgate their own perspectives. Freud profoundly influenced other important thinkers such as Erik Erikson (1963), Erich Fromm (1973), so-called neoanalytic theoreticians (Erikson's approach is sometimes called Ego Psychology because of its emphasis on the development of ego function throughout life) as well as object relations theorists. Object relations theorists such as Melanie Klein (1964) and D. W. Winnicott (1965) altered Freud's original approach by emphasizing the role of relationships with important people (objects), deemphasizing the biological instinctual forces that intrigued Freud, and sometimes highlighting the central role of very early (preverbal) interpersonal relations.

Applications of Psychoanalytic Theory

The most lasting of Freud's contributions may well be his thoughts on resolving psychological problems by talking through them. This notion of talk therapy as potentially curative is so generally accepted in our culture today that it is easy to forget Freud's contribution to our understanding of talk therapy. By listening to his neurotic patients, Freud came to believe that the taproot of their symptoms was buried in unconscious conflicts deriving from early childhood experiences. Lasting relief would come only via gaining *insight* into these dramas and slowly, over a long period of time, working through self-defeating behaviors and feelings. Freud thought the psychoanalyst (a term referring to a psychotherapist with a psychoanalytic orientation) is like an archeologist of the mind (Gay, 1988): The analyst must slowly and meticulously uncover the conflict, utilizing techniques such as dream interpretation, free association, and analysis of the transference with the aim of bringing to light the relic of psychological conflict that is the root of the symptoms. This notion of uncovering the root problem is also widely accepted in our culture today.

Research Methods and Evidence

Freud, as well as many adherents to the psychoanalytic approach, relied on detailed case studies as his preferred method of investigation. In fact, Freud's theory was built almost exclusively upon a limited number of published case studies as well as his own self-analysis. Although case studies provide a wealth of interesting details and may help generate hypotheses for more controlled investigation, it is scientifically risky to rely on case analyses to uncover cause-effect relationships. The limitations of case studies are well known and include a lack of internal validity, difficulty in making generalizations from the case to the population, and the potential biases of the investigator. Some recent approaches to studying psychoanalytic constructs adopt a more controlled empirical methodology. Some of these investigations are far removed from the central ideas of psychoanalytic theory (Dijksterhuis & Nordgren, 2006), whereas others closely parallel it (Epstein, 1994; Westen, 1998).

Evaluation

Many of the theoretical ideas in Freud's body of work are difficult to investigate empirically. Part of this difficulty resides in the imprecise vocabulary of psychoanalytic constructs. Freud did not offer definitions that would lead to careful scientific investigation of his ideas, nor is the theory structured in a way that lends itself to being falsified. The empirical validity of psychoanalytic theory, therefore, is highly questionable. Indeed some (e.g., Macmillan, 1991) argue that Freud's thinking and research methodologies are so profoundly flawed that we should no longer teach his ideas in psychology.

Other Freudian ideas that have been systematically investigated with the tools of modern psychological science have been found to be false or open to significant question at best. For example, consider repression, one of the more important constructs in Freud's thinking. Many scholars (e.g., Loftus & Ketcham, 1994; McNally et al., 2006) have suggested that rather than repress traumatic events, many people are in fact not able to forget them, even though they might want to! Therefore, one of Freud's central tenets about personality may not hold up very well under the scrutiny of controlled investigation.

We should not be overly harsh, however, when evaluating the totality of Freud's work and the impact his ideas continue to have on contemporary psychological science in a number of domains (Westen, 1998). When considering his approach in the light of his historical context, it is easier to see that in his mind, Freud (1928/1961) was adopting a scientific approach. In his book *The Future of an Illusion*, in which he argued that humans should look to science for answers to life's questions rather than to religion, Freud asserted, "No, our science is no illusion. But an illusion it would be to suppose that what science cannot give us we can get elsewhere" (p. 71). Did Freud make this statement merely to convince readers that he was indeed scientific in his approach? Doubtful. Freud himself believed he was adopting a scientific attitude and methodology in his scholarly work. This claim does not excuse the errors Freud made in aspects of his thinking; it simply offers a more nuanced evaluation of his theoretical contributions.

We may also assess the value of Freud's theory utilizing other evaluative criteria. For example, psychoanalytic theory has stimulated vast amounts of empirical research, thus fulfilling an important heuristic function. Freud's theory is also comprehensive and useful, though not parsimonious.

SKINNER'S LEGACY: THE BEHAVIORAL TRADITION

A vastly different but equally controversial approach to understanding personality is the behavioral tradition articulated in its purest form in the work of B. F. Skinner. Elegant in its parsimony, Skinner's operant conditioning perspective continues to generate thoughtful debate, important applications, and continued research. Like Freud, Skinner (1948) hoped his way of thinking would eventually hold sway and transform the field of psychology. Indeed, he believed his ideas could be used to change the very fabric of society.

Important Theoretical Constructs

Understanding that a person is a complex organism behaving in lawful ways, the task of the Skinnerian personality psychologist is to discover the general scientific laws that govern and therefore allow prediction of behavior. Skinner's approach to understanding personality may feel alien because we are so accustomed to looking within the person for the motivations of behavior. However, Skinner (1953, 1974) stridently resisted inner forces as explanations for behavior. Instead, he believed that any presumed internal explanations for behavior were merely explanatory or redundant fictions that one day we would abandon, just as other sciences have abandoned unverifiable causes of events. Any reference to internal states as *explanations* created intellectual traps, Skinner believed, and therefore should be avoided. Although persons might indeed *feel something* as they behave, feelings should not be used as *causal explanations* of behavior. Skinner did not deny the existence of inner states, as is sometimes believed (Debell & Harless, 1992); rather, he rejected using inner states as explanations of behavior.

An example may help clarify this important dimension to Skinner's thinking. John is a 10-year-old boy who frequently fights with his classmates. Viewing John's behavior through a psychoanalytic lens would encourage us to invoke internal motives for his behavior. We might posit, for instance, that John's id is not being adequately restrained by the forces of his ego or superego. Perhaps John has repressed anger toward his father that he displaces onto his classmates so as to get partial gratification of the unconscious hostile impulses he feels toward his father.

To Skinner, John's behavior can best be understood by looking to important environmental contingencies in the form of reinforcers and punishers. For example, perhaps in his home John is able to get his needs met only if he aggressively asserts himself. Generalizing from his home context to the school environment, John may behave similarly at school to the degree that the same behaviors are reinforced in the school context.

Skinner's theory is devoid of constructs. What mattered to Skinner were observable environmental events and the behaviors that derive from those events. Such contingencies of reinforcement and punishment are relevant causal factors. Though biological forces and genetic predispositions place constraints on behavior (e.g., you cannot teach a person to fly no matter how much you reinforce approximations of this behavior), Skinner tended to ignore these influences because they cannot be controlled.

In reading this brief outline of Skinner's (1971) way of viewing personality, it may be apparent that Skinner, like Freud, was a strict determinist. Both Freud and Skinner shared the belief that we are not free to choose our own behavior. Such freedom is more of an illusion than a reality. Unlike Freud, however, Skinner looked to the environment for the determinants of behavior, rather than to internal explanations.

Applications of Behavioral Theory

The applications of Skinner's perspective are numerous and important. The wide array of behavioral modification strategies is a testimonial to the power of Skinner's vision (Kazdin, 2000). For instance, Skinner's ideas have been used to provide behavioral treatment for autistic children; emotionally and behaviorally disturbed children; persons with phobias, sleep disorders, sexual dysfunctions, depression, and other severe mental illnesses; problems with assertiveness; sports-related activities, and so on. As one would expect, all of these approaches target specific behaviors and adopt an empirical approach to evaluating the effectiveness of the intervention.

Research Methods and Evidence

Skinner has been unfairly criticized for viewing humans as "robots" and for dehumanizing them. And yet, his methods of studying behavior argue against this simplistic way of understanding his thinking. A champion of single-subject designs, Skinner was highly concerned with the uniqueness of individuals, an emphasis that is observed in the validation of behavioral treatment approaches.

In general, the behavioral approaches, including those with a cognitive component, rely on well-controlled scientific studies to establish the empirical validity of their propositions and their behavior change interventions. There is strong evidence for the efficacy of behavioral approaches in the domains mentioned (Kazdin, 2000). In fact, the strength of this approach is its hardheaded empirical approach to studying its assertions. If an investigation into a therapeutic strategy does not lend support to the usefulness of the approach, behaviorists will refine the approach and study it further.

Evaluation

If stimulation of debate and research is one measure of a theory's importance, then, like Freud, Skinner's view of human personality is extremely important. In addition, there is a great deal of empirical evidence that the mechanisms underscored by Skinner are indeed powerful shapers of personality. It is fair to say therefore that Skinner's theoretical approach is comprehensive, has empirical validity, and has served an important heuristic function. Perhaps the most notable strength of his thinking, however, is its parsimony.

Despite these strengths, Skinner's approach is not without its problems. Is it legitimate, for example, to describe personality as a complex collection of behaviors? Doesn't this conceptualization leave out something central to what it means to be human? Isn't it too reductionistic? These questions are precisely those that created opposition to his thinking and generated approaches that hoped to recapture the humanity that critics believed was lost. To do so would require again looking inside persons for the motivations of their behavior.

ROGERS' AND MASLOW'S LEGACY: THE HUMANISTIC-PHENOMENOLOGICAL TRADITION

Let's pause for a moment and briefly consider the historical landscape that led to the humanistic-phenomenological tradition. The psychoanalytic approach is sometimes referred to as the "first force" in personality psychology. The second force is the behavioral tradition most carefully elucidated by Skinner. Behaviorism developed in direct opposition to the Freudian tradition. By the 1950s, both of these forces were firmly established in the intellectual environment of American psychology. But opposition to these two forces began to mount, primarily in the work of Carl Rogers and Abraham Maslow. Like Freud, both of these theorists were clinically oriented (though Maslow's graduate training was in experimental psychology), but both strongly opposed some of Freud's and Skinner's central ideas and assumptions about human behavior. Both Rogers and Maslow, for instance, rejected the determinism inherent in the thinking of Freud and Skinner. Instead, the humanistic approach underscores the freedom of individuals to choose their own pathways in life. In addition, Rogers and Maslow assumed that human nature was either inherently good or at least neutral, thus rejecting Freud's claim that humans are driven by primitive, self-serving, and animalistic instinctual forces. Finally, Rogers and Maslow emphasized psychological development throughout the lifespan and, in particular, were optimistic that this development could be healthy. In this sense, Rogers and Maslow anticipated the contemporary Positive Psychology movement (Seligman & Csikszentmihalyi, 2000).

Important Theoretical Constructs

Rogers's (1942) first significant book was *Counseling and Psychotherapy: New Concepts in Practice*, which outlined many of the central features of his theoretical approach. At birth, the organism is the basic psychological reality that, over time, gives rise to more complex personality structures. One of these structures is the self, which roughly corresponds to one's identity. As the self is exposed to conditions of worth, circumstances in which it is valued for certain behaviors or attitudes and devalued for others, it begins to change in order to conform to the conditions under which it is valued. For example, the developing child might grow up in an environment where the direct expression of anger is prohibited. In early life, whenever the self encounters rejection of angry displays, it learns that it is devalued when anger is expressed. Because all persons possess a need for positive regard, they learn to deny or distort the rejected dimension of self. As development proceeds in this fashion, aspects of the person's true nature are denied or distorted. Being shaped by conditions of worth creates a person that is in some ways only a shadow of his or her original nature, so to speak. Overcoming these personal limitations is the task of growing up and becoming more fully functioning. Rogers (1957) was confident that given the proper psychological environment in the formality of a therapeutic relationship or in their everyday lives, people could grow and develop into psychologically healthy individuals.

Maslow shared with Rogers certain fundamental assumptions about personality, but offered unique contributions. Most familiar to many students of psychology is Maslow's (1999) hierarchy of needs, in which he outlined important developmental aspects of his thinking. The essential idea in the hierarchy is that humans progress through phases or levels of development in which lower, more basic needs must be met before the person can begin to satisfy higher needs in the hierarchy. In order, these needs are physiological, safety and security, belongingness and love, self-esteem, and finally self-actualization. Importantly, the lower the need is in the hierarchy, the more powerful it is, and the higher the need, the weaker but more distinctly human it is. The implication of this model is that people are free to devote time and energy to self-actualization only after more basic needs are satisfied.

A more complete understanding of Maslow's thought requires attention to the nature of self-actualization. His thoughts on this subject are what make him stand out as a significant personality theorist. His interest in this important dimension of human psychological development began with his doctoral dissertation. Working with Harry Harlow at the University of Wisconsin, Maslow's dissertation investigated dominance hierarchies in monkeys. Thus, from early in his career, Maslow was interested in the best a species had to offer, so to speak. Extending this interest to humans, Maslow embarked on a research program in which he studied some of the most important figures in human history. For example, he examined the lives of Ralph Waldo Emerson, George Washington Carver, William James, Harriet Tubman, and others, with the aim of understanding the personal characteristics that set these persons apart from normal, everyday people.

Both Rogers and Maslow, therefore, attempted to illuminate the degree to which humans could overcome challenges and live fulfilling, creative, vibrant lives. Like Freud and Skinner before them, the force of their thinking and personalities stimulated others to follow in their footsteps.

The theoretical and empirical work of Rogers and Maslow influenced other, similar approaches. Most of these approaches have an existential orientation in that they consider fundamental questions about the nature of human existence and suffering and emphasize the role of personal choice in directing one's own life to create meaning for it. Less optimistic in their orientations, theorists such as Rollo May (1953) and Viktor Frankl (1959) provide variations on the phenomenological themes developed by Rogers and Maslow.

Applications of Humanistic-Phenomenological Theory

Rogers provided extensive guidance for applying his theoretical constructs to helping persons with psychological problems. In fact, Rogers's work on therapeutic interventions revolutionized how therapists have thought about helping troubled persons. In its purest form, Rogers's belief was that even everyday relationships could be psychologically healing. Whereas psychoanalysis relied heavily on the *techniques* of doing therapy, the Rogerian approach relied on the *person* of the therapist as the healing component. Over the course of his writing, Rogers first referred to his therapy as client-centered (to contrast with the psychoanalytic idea of the patient needing the expert help of the analyst) and then later as person-centered, emphasizing even more the egalitarian relationship between the helper and the person being helped.

What is it about the nature of the therapeutic relationship that is healing? Rogers (1957) asserted there were three *necessary and sufficient* conditions that would inevitably lead to positive therapeutic change. These conditions were that the therapist must (a) show unconditional positive regard (and thus reverse conditions of worth), (b) be congruent or genuine (and thus allow clients to begin to become attuned to their own authenticity), and (c) show empathy toward the client (and thus allow clients to more deeply experience their own feelings). By demonstrating these conditions, Rogers believed clients would begin to move toward a greater understanding of themselves and the possible solutions to existing problems. The role of the therapist is not the expert with answers, but the facilitator to help the client discover his or her own answers. In essence, the goal of Rogerian therapy is for the person to become more self-directing and fully functioning.

Research Methods and Evidence

Unfortunately, Rogers has sometimes been mistakenly regarded as scientifically softheaded in his understanding of personality psychology. This criticism, however, is erroneous. Clearly, some of Rogers' constructs are ill defined and difficult to study empirically. However, Rogers systematically studied many of his central assertions regarding behavior change in psychotherapy. Utilizing a method known as the Q-sort, Rogers (1961) investigated how self-concept changes over time in counseling. In addition, he was the first to bring audiotape technology into therapy sessions so that these sessions could be recorded and later subject to empirical scrutiny. This practice is still common today. For these efforts and others, in 1956 Rogers was awarded the first Distinguished Scientific Contribution Award by the APA.

Evaluation

Like the psychoanalytic and behavioral traditions, the perspectives of Rogers and Maslow are significant. Rogers's theory is more carefully articulated and developed than Maslow's, but both viewpoints are comprehensive, internally consistent, and have stimulated a good deal of research and debate. Neither is very parsimonious. It is fair to criticize them on the basis of not clearly operationalizing important constructs. Like Freud, both these theorists did not specify the conditions under which their ideas could be proved false. It is much easier to explain behavior after the fact, rather than predict how behavior might unfold in the future.

CONTRIBUTIONS OF COGNITIVE THEORISTS

Cognitive approaches to understanding personality are in some ways similar to the behavioral tradition because they share a commitment to the empirical validation of ideas. If fact, some of the theorists in this tradition, such as Albert Bandura, are paragons of careful scientific analysis. The important contrast with the radical behavioral approach of Skinner (1990) is that cognitions are internal events that are seen as causal explanations of behavior, something that Skinner rejected throughout his entire career (Skinner, 1990).

Important Theoretical Constructs

Let's start our consideration of this set of theories with a perspective that is in some ways difficult to categorize. The Personal Construct Theory of George Kelly (1955) was an important precursor to other cognitive theories. In a nutshell (keep in mind that Kelly's 1955 tome had 1,200 pages!), Kelly asserted that people are like scientists who possess a set of personal constructs or cognitive frameworks that they use to interpret their experiences. Personal Construct Theory has one foot in the cognitive tradition and one in the subjective experiences or interpretations of the phenomenological viewpoint. These personal constructs both derive from past experience and guide a person's interpretation of ongoing experience. One's belief system or set of personal constructs is what is most important, not the particular events of one's life. For instance, if a child loses a parent, the child might develop a system of constructs that construes relationships as always ending in hurt or loss.

This type of belief system will have an obvious impact on how this developing person behaves in future relationships. On the other hand, the child might grow into an adult who values every moment of his or her relationships due to the knowledge of the potential brevity of relationships. In either scenario, the events are the same; the child's system of personal constructs is what shapes personality.

Albert Bandura (1977) also recognized the power of beliefs in shaping personality. Whereas Skinner was a strict environmental determinist, Bandura adopted a position he called *reciprocal determinism*. Bandura understood that behavior is shaped by environmental factors. He also recognized that behavior impacts the environment and that internal characteristics of the person, such as emotional states or beliefs, also affect behavior. These forces (environment, behavior, and internal person variables) reciprocally determine one another.

Bandura asserted that because people are thinking, information-processing organisms, they can learn by observing the behavior of others and can modify their own behavior by vicariously experiencing the consequences that others experience. If you are a student in a classroom and observe that another student is ridiculed by the professor for asking a question, your probability of asking a question in that class is diminished. The unpleasant consequence of ridicule did not happen directly to you, but your own behavior was modified as a result of your observations of the other person, the so-called model. Bandura (1977) labeled this type of learning "observational learning," and it was a central construct in the initial conceptualization of his theory known as *social learning theory*. Bandura showed that a person could learn something (create an internal representation of a behavior) without ever actually performing the behavior. He demonstrated this principle in his now-famous Bobo doll study (Bandura, 1965).

Over time, Bandura (1986, 2006) more clearly outlined the set of cognitive processes that allow for persons to be self-regulators of their own behavior, which led to a shift from describing his theory as social learning to a social cognitive perspective. In more recent years, Bandura (1997) carefully explicated an important cognitive system known as *self-efficacy*, the *belief* that one can carry out behaviors that will lead to desired outcomes. Self-efficacy beliefs are important causal factors in driving behavior in a variety of domains (e.g., academic, health, athletic), something that has important implications for behavioral change applications.

Applications of Cognitive Theory

A variety of applications grow out of theories within the cognitive domain. Though not comprehensive theories in their own right, many scholars have taken the fundamental principles of the cognitive approach and applied them to systems for helping people.

Aaron Beck's (1976) well-known cognitive approach to treating depression and other problematic emotional states is a significant application of these ideas. Donald Meichenbaum's (1993) Stress Inoculation training is another application for helping people develop effective strategies for dealing with stressful situations. In addition, Albert Ellis (1995) offered his Rational Emotive Behavior Therapy as a powerful means of changing behavior via the alteration of irrational belief systems.

Evaluation

As in the behavioral approach, those in the cognitive arena have relied on systematic investigation of their claims via well-controlled scientific studies. In addition, both Bandura and Kelly offered comprehensive, empirically sound theoretical perspectives. Though not parsimonious, these theorists attempted to strip their perspectives of unnecessary constructs, though their final solutions are complex. Bandura's approach has received more widespread attention and has therefore served an important heuristic function.

TYPES, TRAITS, AND THE BIOLOGICAL UNDERPINNINGS OF PERSONALITY

Type and trait approaches to understanding personality have a long and distinguished history. Although I have saved this area of personality psychology for the latter part of the chapter, this approach is in some ways the oldest perspective in the study of personality. Early typologies of personality can be traced back thousands of years to the Greek physician Hippocrates. He maintained that the human body consisted of four humors or biochemical systems known as blood, black bile, yellow bile, and phlegm. Several hundred years later, the 2nd-century Roman physician Galen asserted that an imbalance in these humors would create personality types corresponding to sanguine (cheerful), melancholic (sad), choleric (angry and hostile), and phlegmatic (apathetic; Kagan, 1994). Thus, a personality system was born.

Interestingly, modern techniques of neuroscience have borne out some of the observations of these ancient physicians. In fact, developments in our understanding of the biological and genetic underpinnings of personality, along with advances in statistical techniques, have led to a resurgence in this way of understanding personality. More than the other viewpoints we have examined, this group of perspectives underscores the ways nature and nurture work together to create individual differences in personality.

Important Theoretical Constructs

The two dominant constructs in this area are traits and types. A trait is an enduring personality characteristic, which is presumed to have some stability and will explain various dimensions of behavior. Using traits, we slice personality into component parts and use them as explanations for individual differences. For example, if one student frequently speaks up in class but another remains silent, we might attribute this behavioral difference to the relative strength of the trait of extraversion in these two persons, with a high degree in the former and a low one in the latter. Traits are therefore continuous variables that vary in amount from person to person.

A type is a general category used to describe a person. Rather than slicing personality into component parts, types sort people into groups. Utilizing the example above and adopting a type approach, we might conclude that the former student corresponds to the category of extravert, whereas the latter person is an introvert. Types are therefore discontinuous variables; one is either an extravert or an introvert, but not both. One of the most well-known typologies of personality is the Type A, Type B distinction, first described by Freidman and Rosenman (1974), two cardiologists who observed that certain behavior patterns seemed to predict heart disease. Following their intuitions, they empirically investigated their hunches by studying a group of middle-aged men over an eight-year period and uncovered the link between traits such as impatience, hostility, driven competitiveness, and heart disease. Further studies highlighted the importance of *hostility* as the toxic dimension in the Type A pattern (Friedman & Booth-Kewley, 1987; Greenglass & Julkunen, 1991).

Raymond Cattell (1952, 1965) was a groundbreaking trait theorist who utilized the statistical technique of factor analysis to investigate the dominant traits in human personality. Referring to these traits as factors, Cattell believed that 16 source traits could best account for individual differences in personality. The 16 Personality Factor questionnaire, or 16 PF for short, was the instrument Cattell developed for assessing these factors.

To his credit, Cattell recognized that more than just questionnaire data should be used in understanding the uniqueness of a person. To supplement questionnaire or Q-data, Cattell argued that psychologists should also obtain L-data, or information gathered from the person's life. For example, L-data might be derived from school records, job histories, social history, and so on. T-data or test data correspond to information gathered from real-life or experimentally contrived situations in which the person's behavior can be observed. For instance, you can observe someone's behavior in a variety of naturally occurring social situations or, if you are a psychological researcher, in a situation that is part of a psychological study. Cattell's selection of the term T- or test-data is confusing (Funder, 2004) because most persons think of questionnaires as "tests," but T-data should be seen as distinct from Q-data.

As it turns out, I have oversimplified the distinction between traits and types, as can be seen in the work of Hans Eysenck (1967), an influential scholar who has intertwined the ideas of types, traits, and the underlying biological mechanisms that shape personality. Believing that the major aspects of personality are largely genetically determined, Eysenck updated our understanding of personality types. He believed that types are continuous variables along which people differ, though most people possess an average amount of the type dimension. Utilizing the factor analytic approach, Eysenck (1967) believed that three types account for the greatest degree of individual differences in personality. These types are Extraversion, Neuroticism, and Psychoticism, which are measured using the Eysenck Personality Questionnaire. Eysenck also claimed that these basic personality dimensions are undergirded by inherited differences in nervous system reactivity, the details of which go beyond the scope of this chapter.

At present, the most dominant contemporary theory in the trait tradition strikes a balance between the 16 factors in Cattell's model and the 3 in Eysenck's. McCrae and Costa (1987) are the primary architects of the perspective known as the Big Five Theory or the Five Robust Factors, though their work built on the foundation of Norman (1963). Like other trait theorists before them, McCrae and Costa utilized a factor analytic approach and asserted that five dimensions capture the essence of human personality: neuroticism, extraversion, openness, agreeableness, and conscientiousness. Big Five theorists avow that these "supertraits" are independent of one another and can describe personality across cultures, though there may be differences in relative strength across cultures, as well as sex differences within cultures (Paunonen, 2003).

I have outlined the dominant perspectives in the type and trait domain. All these approaches owe a debt, however, to Gordon Allport (1937), who years ago attempted to document the range of trait vocabulary used in describing personality. More complex than developing a taxonomy of traits, Allport's (1938, 1955) contributions should not be forgotten, although they are less influential today.

Applications of Trait and Biological Theory

Compared to other perspectives we have examined, the type and trait approaches are relatively silent regarding applications that may lead to behavior change. Because most types and traits are seen as having an important genetic and biological component, behavioral change is difficult, though not impossible. A frequently used metaphor in this tradition is to declare that personality is akin to plaster, not granite. It is difficult to alter, though not impossible. Adopting a similar viewpoint, Kagan (1994) argued that certain personality characteristics are largely inherited, although "biology is not destiny." For example, if a child is born with a genetic predisposition to be introverted, parents can take steps to help this child deal more effectively with novel or potentially intimidating situations. Through these experiences, the child may learn to be bolder when encountering situations that might cause him or her to withdraw. At the same time, although biology is not destiny, it does place parameters on the range of behavioral change that is possible (Kagan, 1994).

Research Methods and Evidence

Type and trait theorists have adopted a strong empirical tradition, often relying on a procedure known as *factor analysis* (Cattell, 1952). As a data-reduction strategy, factor analysis is a powerful statistical procedure that seeks to find broader, underlying structures (factors) that can account for a greater number of more specific characteristics. Although this approach is used widely in a number of psychological studies, both inside and outside the personality psychology domain, it is not without its critics. Skeptics argue, for instance, that factor analysis suffers from the "garbage in, garbage out" problem. A factor solution can be only as good as the data that are originally entered into the factor analytic study. If garbage goes in at the start, it is inevitable that garbage comes out at the end, albeit rearranged and more neatly packaged. In addition, the results of a factor analytic study merely uncover the underlying factors; they do not identify or name those factors. The naming of the factors is left to the informed but subjective opinion of the investigator.

Evaluation

A great deal of empirical evidence has been garnered in support of the type and trait tradition. The scholars in this arena should be applauded for their painstaking efforts to find empirical support for their claims. Moreover, these approaches tend to be comprehensive in their scope. More parsimonious than Freud, these theories lack the simplicity found in the behavioral tradition. A relative weakness is their lack of specificity regarding mechanisms of personality change, although advances in biochemical interventions (e.g., psychotropic medications) may be a burgeoning avenue for the study of behavioral or personality change.

One final evaluative comment is important and is something that may have already occurred to you. Each of the perspectives we have considered in this section (Cattell, Eysenck, McCrae, and Costa) all claim to have found the basic elements of personality, and all have used a strong empirical approach to developing their theories. Yet each theorist has arrived at a different number and set of basic elements, although there is some overlap in their thinking (e.g., extraversion). The resolution of this conundrum will depend on further empirical work.

SUMMARY

The field of personality psychology is one of the most stimulating areas in psychology. It weaves together aspects of individual differences and commonalities. It draws our attention to the power of nurture…*and* to the hidden authority of our unique genetic code. It compels us to consider important developmental questions such as the continuity and discontinuity of our identities. All of the perspectives presented in this chapter are in a sense competing for superiority in their explanatory power. Yet none of them is dominant as an overarching explanatory system the way philosopher of science Thomas Kuhn (1970) envisioned. The ongoing intellectual and empirical debate regarding which approach is best remains undecided, with no particular perspective emerging as *the* paradigm to follow. In part, this lack of resolution makes this field dynamic and exhilarating!

I would like to end this chapter with two compelling questions. The first is, "Why do many of these theories persist in the absence of solid empirical evidence?" This question has been raised by Monte (1999), who provided an insightful answer. These theories persist, Monte argued, because they deal with the important questions of human existence, the so-called "big potato" questions. They ask us to think about questions such as *What makes us human? What distinguishes one person from another? To what extent are we free to choose our own destinies? Is there hope for a world marked by peace?* These questions and others like them are extremely difficult to answer, but they compel personality psychologists to continue to probe for solutions.

The second question is, "Do these theories really capture what it means to be a person?" As I noted at the outset of this chapter, we are all implicit personality theorists who are constantly sizing up people and wondering what makes them tick. When we formalize our thinking into explicit, testable theories, do we really capture the essence of individual lives? Is there something inevitably missing when we attempt to do so? As Monte (1999) observed, the person in a scientifically verifiable theory may not be the person who sits across from you in your favorite coffee shop. Herein lays a central problem in the formal theories of personality psychology. At the same time, this fundamental quandary creates ambiguities and gray areas, the very stuff that ignites continued thought and investigation.

REFERENCES AND FURTHER READINGS

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart & Winston.

Allport, G. W. (1938). Personality: A problem for science or a problem for art? *Revista de Psihologie, 1*, 1–15.

Allport, G. W. (1955). *Becoming: Basic considerations for a psychology of personality*. New Haven, CT: Yale University Press.

American Psychological Association. (2007). *APA dictionary of psychology*. Washington, DC: Author.

Anderson, D. D., Rosenfeld, P., & Cruikshank, L. (1994). An exercise for explicating and critiquing students' implicit personality theories. *Teaching of Psychology, 21*, 174–177.

Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology, 1*, 589–595.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*, 164–180.

Beck, A. T. (1976). *Cognitive therapy for the emotional disorders*. New York: International Universities Press.

Breuer, J., & Freud, S. (1895). *Studies on hysteria*. New York: Basic Books.

Cattell, R. B. (1952). *Factor analysis*. New York: Harper.

Cattell, R. B. (1965). *The scientific analysis of personality*. Baltimore: Penguin.

Cohen, D. B. (1999). *Stranger in the nest: Do parents really shape their child's personality, intelligence, or character?* New York: Wiley.

Debell, C. S., & Harless, D. K. (1992). B. F. Skinner: Myth and misperception. *Teaching of Psychology, 19*, 68–73.

Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1*, 95–109.

Ellis, A. (1995). Rational emotive behavior therapy. In R. J. Corsini & D. Wedding (Eds.), *Current Psychotherapies* (5th ed, pp. 162–196). Itasca, IL: Peacock.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709, 724.

Erikson, E. H. (1963). *Childhood and society* (2nd ed.). New York: Norton.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Charles C. Thomas.

Frankl, V. E. (1959). *Man's search for meaning*. New York: Simon & Schuster.

Freud, S. (1961). *The future of an illusion*. New York: Norton. (Original work published 1928)

Freud, S. (1977). *Five lectures on psycho-analysis*. New York: Norton. (Original work published 1909)

Friedman, H. S., & Booth-Kewley, S. (1987). Personality, Type A behavior and coronary heart disease: The role of emotional expression. *Journal of Personality and Social Psychology, 53*, 783–792.

Friedman, H. S., & Rosenman, R. H. (1974). *Type A behavior and your heart*. London: Wildwood House.

Fromm, E. (1973). *The anatomy of human destructiveness*. New York: Holt, Reinhart & Winston.

Funder, D. C. (2004). *The personality puzzle* (3rd ed.). New York: Norton.

Gay, P. (1988). *Freud: A life for our time*. New York: Norton.

Greenglass, E. R., & Julkunen, J. (1991). Cook-Medley hostility, anger, and the Type A behavior pattern in Finland. *Psychological Reports, 68*, 1059–1066.

Hall, C. S. (1954). *A primer of Freudian psychology*. New York: The New American Library.

Hall, C. S., & Nordby, V. J. (1973). *A primer of Jungian psychology*. New York: The New American Library.

Hamer, D., & Copeland, P. (1998). *Living with our genes*. New York: Anchor.

Harris, J. R. (1998). *The nurture assumption: Why children turn out the way they do*. New York: The Free Press.

Hopcke, R. H. (1992). *A guided tour of the collected works of C. G. Jung*. Boston: Shambhala.

Jung, C. G. (1961). *Memories, dreams, reflections*. New York: Random House.

Kagan, J. (1994). *Galen's prophecy: Temperament in human nature*. New York: Basic Books.

Kazdin, A. E. (2000). *Behavior modification in applied settings* (6th ed.). Belmont, CA: Wadsworth.

Kelly, G. A. (1955). *The psychology of personal constructs* (Vols. 1 & 2). New York: Norton.

Klein, M. (1964). *Contributions to psychoanalysis, 1921–1945*. New York: McGraw-Hill.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: The University of Chicago Press.

Loftus, E., & Ketcham, K. (1994). *The myth of repressed memory: False memories and allegations of sexual abuse*. New York: St. Martin's Press.

Macmillan, M. (1991). *Freud evaluated: The completed arc.* Amsterdam: North Holland.

Maslow, A. (1999). *Toward a psychology of being* (3rd ed.). New York: Wiley.

May, R. (1953). *Man's search for himself*. New York: Norton.

McCrae, R. R., & Costa, P. T. (1987). Validation of the Five Factor Model across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81–90.

McNally, R. J., Clancy, S. A., Barrett, H. M., Parker, H. A., Ristuccia, C. S., & Perlman, C. A. (2006). Autobiographical memory specificity in adults reporting repressed, recovered, or continuous memories of childhood sexual abuse. *Cognition & Emotion, 20*, 527–535

Meichenbaum, D. (1993). Stress inoculation training: A 20-year update. In P. M. Lehrer & R. L. Woolfolk (Eds.), *Principles and practice of stress management* (2nd ed., pp. 373–406). New York: Guilford.

Miller, A. (1981). *The drama of the gifted child*. New York: Basic Books.

Monte, C. F. (1999). *Beneath the mask: An introduction to theories of personality* (6th ed.). Fort Worth, TX: Harcourt Brace.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66*, 574–583.

Paunonen, S. V. (2003). Big Five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology, 84*, 411–422.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Rogers, C. R. (1942). *Counseling and psychotherapy: New concepts in practice*. Boston: Houghton Mifflin.

Rogers, C. R. (1951). *Client-centered therapy*. Boston: Houghton Mifflin.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95–103.

Rogers, C. R. (1961). *On becoming a person*. Boston: Houghton Mifflin.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14.

Skinner, B. F. (1948). *Walden two.* New York: Macmillan.

Skinner, B. F. (1953). *Science and human behavior*. New York: The Free Press.

Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Knopf.

Skinner, B. F. (1974). *About behaviorism.* New York: Knopf.

Skinner, B. F. (1990). Can psychology be a science of mind? *American Psychologist, 45,* 1206–1210.

Sulloway, F. J. (1979). *Freud, biologist of the mind.* New York: Basic Books.

Westen, D. (1998). The scientific legacy of Sigmund Freud: Toward a psychodynamically informed psychological science. *Psychological Bulletin, 124,* 333–371.

Winnicott, D. W. (1965). *The maturational process and the facilitating environment*. New York: International Universities Press.

Wordsworth, W. (2002). *Selected poetry of William Wordsworth* (M. Van Doren, Ed.). New York: Modern Library. (Original work published 1802)

44

INTELLIGENCE

JOSEPH J. RYAN
University of Central Missouri

For centuries, persons making significant contributions to science, industry, art, politics, or even criminal pursuits were viewed as intelligent, whereas those who failed in school, at work, or as a member of society were thought to represent the lower end of the ability distribution. The ancient Chinese as well as the ancient Greeks recognized the importance of individual differences in ability for the selection of soldiers and civil servants. As part of modern Western culture, children and adults routinely test their abilities (intelligence) against peers via school grades, extracurricular activities (e.g., debate clubs), games of skill (e.g., chess), and even party games (e.g., Trivial Pursuit) and television shows (e.g., Jeopardy). Although the subjective assessment of intelligence has been going on since the dawn of civilization and will continue in perpetuity, the formal measurement of the construct dates to the latter part of the 19th century. This chapter addresses the topic of human intelligence by expounding on its history, contemporary explanatory theories, controversial issues, practical applications, and future trends in the field.

HISTORY

From the time of Aristotle, philosophers and academics have debated and discussed the construct of human intelligence. The ancient Greeks and Chinese, for example, were interested in identifying highly "intelligent" individuals who could learn quickly, benefit from formal education, and function successfully within complex social and military systems. Around 2200 BCE, Chinese emperors were actually using competitive written civil service examinations to select employees for their government bureaucracies. Although attempts to identify differences in ability among people have been taking place for thousands of years, the scientific study of intelligence is a relatively recent event, one that began in the latter part of the 1800s.

Sir Frances Galton, a British scientist and scholar, believed intelligence was mainly inherited and that the human race could be improved by selectively breeding the healthiest, brightest, and most physically capable individuals. In order to identify the "best" people, he established the first anthropometric laboratory in 1884. Galton assumed that he could use tests of physical and sensory ability to measure intelligence. He reasoned that if information about the world comes to us via sensory processes, then more intellectually capable persons will perform better than less capable people on separate tasks of reaction time, strength of movement, visual discrimination, kinesthetic discrimination, and auditory discrimination.

James McKeen Cattell, who earned his PhD at the University of Leipzig under Wilhelm Wundt in 1886, shared Galton's interest in identifying individual differences using sensory-motor tasks, which he believed measured intellectual functions. After completing postdoctoral studies under Galton, he returned to the United States in 1888 as professor of psychology at the University of Pennsylvania. He immediately founded a psychological laboratory at Penn, began collecting sensory-motor data on healthy volunteers, and coined the term "mental tests" to describe his procedures. In 1891 he moved to Columbia University and, in addition to assuming administrative duties as the department head, continued collecting sensory-motor information on entering freshmen for almost a decade. One of Cattell's graduate students, Clark Wissler, designed his dissertation research to test the hypothesis that sensory-motor tests are valid measures of intellectual ability. Using college student data, Wissler reported that Cattell's sensory-motor tests did not correlate meaningfully with one another, nor did they show significant associations with academic performance. Conversely, academic grades in various subjects (e.g., Latin and mathematics) produced significant intercorrelations. These findings suggested that Cattell's tests lacked validity as measures of intelligence and, in the opinion of some writers (Sokal, 1987), served as the terminal blow to the viability of the anthropometric testing movement.

At the dawn of the 20th century, policies requiring compulsory school attendance for all children were instituted in Western Europe and the United States. To help the Parisian public schools comply with the directive, Alfred Binet and his colleague Theodore Simon were recruited by the French government to develop quantitative techniques for the assessment of school-age children. Unlike Galton and Cattell, who thought that sensory-motor tasks measured intelligence, Binet was convinced that accurate assessment required items tapping higher mental processes (e.g., abstraction, memory, and novel problem solving) in addition to those that measure sensory-motor functions. In 1905 Binet and Simon published a 30-item age-scale that allowed one to determine if an examinee was functioning below, at, or above the level of the average child of the same chronological age. Sattler (2001) describes the 1905 scale as the first practical intelligence test because it provided instructions for administration and was composed of items that were arranged according to difficulty levels. For example, to pass the first item the child followed a light with his eyes, whereas the last question required the examinee to distinguish between the concepts of "sad" and "bored." The original scale was revised and expanded in 1908 (which introduced the concept of mental age, or MA) and again in 1911. The scales were interpreted as follows: mental age ≤ 2 = idiot, 2 to 7 years = imbecile, and ≥ 8 years = moron (Watson, 1963). Henry H. Goddard translated the 1905 and 1908 editions into English and standardized the latter using 2,000 American children. To insure that interpretation of mental age simultaneously considered the examinee's chronological age, William Stern introduced the concept of the *mental quotient* in 1912 (Sattler, 2001). This value is calculated by dividing MA by chronological age (CA) to reflect overall intelligence. For example, a seven-year-old child with an MA of nine has a mental quotient of 1.28.

Perhaps one of the most important contributions to the field of mental testing occurred when Lewis Terman, a professor at Stanford University, published the 1916 Stanford Revision of the Binet-Simon scale. Terman expanded upon Binet's ideas, developed new test items, collected normative data on 1,400 children and adults living in California, and converted the *mental quotient* to an *intelligence quotient* or IQ (mental age ÷ chronological x 100). Terman also devised a system for classifying intelligence in terms of IQ. On the American version of the scale, the previously mentioned seven-year-old child with an MA of nine years would have an IQ of 128 (MA/CA x 100), a value falling within the superior range of intellectual functioning. The 1916 revision, which was popularly known as the Stanford-Binet Intelligence Scale, quickly became the standard for intelligence assessment throughout the United States. Although researchers have revised and modified the scale numerous times (1937, 1960, 1972, 1986, and 2003), its popularity began declining in the 1940s and 1950s because of serious limitations when used to evaluate adolescents and adults with a variety of psychiatric or neurological conditions.

David Wechsler (1939) was one of many psychological practitioners who recognized that the Stanford-Binet was unsuitable in a variety of clinical situations. For example, the scale was designed for children and lacked face validity with older examinees. There were also problems with an overemphasis on verbal ability and utilization of the MA construct for calculating IQs. The latter concern was due to the underlying assumption of the 1916 Stanford-Binet that MA ceased to increase after 16 years of age (Kaplan & Saccuzzo, 1997). Thus, an examinee who takes the test with a CA of 16 years and earns an MA of 16 years receives an IQ of 100 (16 ÷ 16 x 100 = 100). However, if the examinee is retested two years later and obtains the same MA, the resulting IQ will show a marked and clearly misleading decline (16 ÷ 18 x 100 = 89). To control for this problem, the highest CA that could be used in calculating IQ values on the 1916 scale was 16 years. Although the 1937 revision of the Stanford-Binet was a much-improved instrument, suitability of the scale for use with adults and problems with reliance on the MA for derivation of ratio IQs remained. It is also noted that the standardization sample for the 1937 revision included only examinees in the age range 18 months to 18 years.

As chief psychologist at Bellevue Hospital in New City, Wechsler was acutely aware of the need for an intelligence measure that corrected the problems associated with the Stanford Binet. Drawing on his extensive background and knowledge of the World War I military testing program, everyday clinical assessment, and the scientific literature, he was able to organize a number of previously published and well-known tests into a composite instrument. His new test, which was called the Wechsler-Bellevue Intelligence Scale-Form I (W-B I; Wechsler, 1939), contains 10 standard subtests and one supplementary subtest that yield separate Verbal Scale and Performance Scale

IQs as well as a global intelligence estimate called the Full Scale IQ. The W-B I is a point scale that groups test items according to content, whereas the fist seven editions of the Binet scales grouped items by age level, regardless of content. To solve the problems associated with the MA and ratio IQ, Wechsler eliminated the former concept entirely and replaced the ratio IQ with the deviation IQ. On the W-B I and subsequent editions of the scale, the deviation IQ involves transforming raw scores to standard scores with a designated mean of 100 and standard deviation of 15. The W-B I standardization sample consisted of 1750 males and females, with raw score-to-IQ conversion tables covering the age ranges 10 years to 59 years on the Verbal and Full Scales and 10 years to 49 years on the Performance Scale.

In the United States, the W-B I rapidly became the preferred instrument for assessing adolescent and adult intelligence. The scale was revised and expanded in 1946, 1955, 1981, and 1997, with the latest edition being designated the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997). The constant process of scale revision also produced a series of tests for children in the age range 6 years to 16 years, including the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), WISC-R, WISC-III , and WISC-IV. For younger children there is the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1967) and its newer derivatives, the WPPSI-R and the WPPSI-III. Recent surveys have demonstrated that in the United States the Wechsler Scales are, for all practical purposes, the most frequently used tests of intelligence for children and adults (Rabin, Bar, & Burton, 2005).

THEORIES OF INTELLIGENCE

The activities just described took place in the absence of a generally accepted definition or theory of intelligence. Although a number of symposia and conferences were held over the years to resolve this situation, to date a consensus has not been achieved. Thus, the following discussion covers four theories of intelligence that this writer feels have the most adherents and would be of interest to the present readership.

Charles E. Spearman (1923) proposed the two-factor theory of intelligence. It is based on the observation that persons who excel on one type of intellectual task (e.g., mathematics) tend also to perform well on others (e.g., defining words). Spearman noted that when individuals were given ability tests tapping a variety of content (e.g., numerical problems and visual spatial designs), the resulting test scores consistently yielded positive intercorrelations. According to the theory, a general factor "*g*" reflects the component that is shared with all conceivable measures of intellectual functioning, and a specific factor "*s*" represents the unique component of each test. The two-factor theory provides a reasonable foundation for the measurement of intelligence with tests such as the Binet and Wechsler because they provide an aggregate or composite MA and/or IQ score that is based on items and subtests of diverse manifest content.

Today some experts consider "*g*" to be a variable accounting for differences in IQ scores across individuals, with each person occupying one of the levels of "*g*" (Borsboom & Dolan, 2006). High "*g*" individuals demonstrate a plethora of generic abilities, including efficient and rapid learning and good reasoning, problem-solving, and conceptual skills. However, IQ scores are not pure measures of "*g*" as evidenced by the fact that the construct accounts for only 50 percent of the variance in the WAIS-III Full Scale IQ (Sattler & Ryan, 2001). Nevertheless, when the IQ is used as a measure of general intelligence it proves to be a highly practical piece of information for predicting favorable outcomes such as job performance, educational attainment, socioeconomic status, and longevity (Gottfredson & Deary, 2004). General intelligence as reflected in IQ scores also predicts unfavorable outcomes, including poverty, welfare use, incarceration, and cause of death (Gottfredson, 2002).

Raymond B. Cattell (1963) postulated that Spearmen's "*g*" could be divided into two separate components, which he labeled *crystallized* and *fluid ability*. The former construct reflects the influence of formal education and acculturation. It represents well-learned facts and cognitive operations and may be assessed with tests of vocabulary, general information, and academic achievement. Individuals can increase their levels of crystallized "*g*" throughout the life span via formal education, personal reading, socialization, and other activities that require new learning and novel problem solving. Conversely, fluid "*g*" is akin to innate ability that reflects the biological integrity of the central nervous system. Fluid ability peaks in adolescence or young adulthood and declines with advancing age and/or neurological disease or insult involving the brain. Fluid ability is not related to education or acculturation and is assessed by tests of working memory, matrices, and concept formation. From a practical standpoint it appears that fluid ability plus motivation begets crystallized ability and achievement. Cattell's theory has been modified and expanded (see Carroll, 1997) and stands today as the most comprehensive and well-validated description of human cognitive abilities. The work of Cattell, Horn, and Carroll, which is today referred to as the Cattell-Horn-Carroll Theory (CHC), has had a profound influence on the field of intelligence assessment. For instance, the WAIS-III and WISC-IV for the first time include a matrix reasoning subtest designed specifically to measure fluid intelligence, and the Woodcock-Johnson Test of Cognitive Abilities-Third Edition, Stanford-Binet Intelligence Scale-V, and Kaufman Assessment Battery for Children-Second Edition each provide summary scores measuring crystallized and fluid intelligence.

A theory of multiple intelligences (MI) was introduced by Howard Gardner in 1983. He posits the existence of at least eight intelligences that are relatively independent of one another and are unrelated to "*g*," with two possible exceptions . The intelligences conceptualized by Gardner are as follows:

1. *Linguistic:* The capacity to use language effectively and to learn new languages. This is characteristic of many groups including attorneys, writers, and poets. Because of its linguistic nature, this domain is likely to show an association with "*g*."
2. *Spatial:* The capacity to visually perceive the world and discriminate patterns. This ability is important for airplane pilots, surgeons, and sculptors, among others.
3. *Logical-Mathematical:* The ability to learn higher mathematics, display logical thinking, and design and carry out scientific experiments. Occupations requiring a good deal of this ability include mathematician, scientist, and logician. This domain is likely to show an association with "*g*."
4. *Interpersonal:* This involves sensitivity to the feelings, motivations, and intentions of other people. This capacity characterizes successful politicians, salespeople, and therapists.
5. *Intrapersonal:* The capacity to understand oneself in terms of personal fears, motivations, and capacities. This ability is demonstrated by persons who are effective self-monitors and who can properly utilize their intellectual assets.
6. *Naturalistic:* This capacity involves the ability to recognize and classify different species and to recognize natural patterns. Successful biologists, farmers, and gardeners demonstrate these abilities.
7. *Musical:* The ability to appreciate music and to demonstrate talent regarding the discrimination of tones in terms of melody, intensity, and rhythmic patterns. Groups with musical intelligence include singers, composers, and musicians.
8. *Body-Kinesthetic:* This refers to the ability to use the whole body to create products or solve problems. Successful surgeons, dancers, and athletes are likely to possess this capacity.

MI theory has been embraced by many professionals within the educational community, partly because of Gardner's unusual approach to assessing intelligence and partly because of his argument that assessments must test examinees' proficiencies in completing culturally valued tasks. For example, in his work with preschool children Gardner has evaluated Logical-Mathematical, Body-Kinesthetic, and Spatial intelligences by having examinees disassemble and reassemble everyday items such as doorknobs and pencil sharpeners (Visser, Ashton, & Vernon, 2006). As interesting as this approach may appear, neither Gardner nor any test publisher has yet to produce a single standardized instrument based on MI theory. Moreover, empirical research has contradicted Gardner's contentions by demonstrating that when tests that appear to measure the various intelligences are simultaneously subjected to statistical analyses, many are substantially intercorrelated and show strong loadings on a "*g*" factor. Visser et al. (2006) imply that it is (a) premature to employ MI theory as a basis of curriculum development and (b) unlikely that MI theory provides any new information about the structure of intellect that is not already contained in the Spearman and CHC models discussed above.

Another model of intelligence is the triarchic theory of Robert J. Sternberg (1985, 1997). This model emphasizes "practical intelligence" along with distinctions among analytic abilities, creative abilities, and practical abilities. Sternberg claims that these constructs differ from "*g*" and predict real-world success as well as or better than do traditional IQ measures. The intelligences in the triarchic theory are as follows:

1. *Analytic*—The ability to solve complex problems, analyze information, think critically, acquire new concepts, and successfully utilize academic knowledge. This ability is similar to what conventional IQ tests purport to measure. This is the componential part of the theory that associates internal mental processes with intelligence.
2. *Creative*—The ability to generate ideas, identify problem-solving options, determine which problems need to be solved, and deal effectively with novelty. This is the experiential part of the theory that is thought to relate intelligence to the individual's internal and external worlds.
3. *Practical*—The ability to deal effectively with real-world problems using tacit knowledge. The latter concept refers to information and skills that reflect "knowing how to do things" within a specific culture or environment. Tacit knowledge is not typically taught in school or explicitly verbalized. This is the contextual dimension of the theory and reflects the application of analytic and creative intelligences to achieve personal goals within a specific environment.

The triarchic theory assumes that "*g*" is related only to analytic abilities and that the analytic, creative, and practical intelligences are independent of one another. However, when researchers administered the Sternberg Triachic Abilities Test (STAT) and a conventional measure of "*g*" to a sample of gifted students, highly significant correlations emerged between "*g*" and the three intelligences, which were also significantly intercorrelated (Brody, 2003). Thus, the three triachic abilities are not independent of one another and "*g*" is a component of each. Although Sternberg's theory is both interesting and highly creative, it would be premature to embrace this approach uncritically. Additional research is needed to demonstrate that the three intelligences represent useful constructs and that they provide information that is not already provided by other theories and methods (Gottfredson, 2003a, 2003b).

CONTROVERSIAL ISSUES

An important concern for anyone interested in the field of human intelligence is the failure of scientists and practitioners to produce a universally accepted definition of the construct. Nevertheless, standardized tests such as the Wechsler Scales and the Stanford-Binet are frequently administered in schools, clinics, and industry and provide useful information about the cognitive functioning of individual examinees, especially when interpreted in conjunction with scores from other tests (e.g., language and memory) and information on educational attainment, social adaptation, and medical status. Because IQ scores carry significant implications for the future of the examinee (e.g., placement in a gifted program) or his/her current situation (e.g., competency to stand trial for a legal offense), it would be helpful to know whether the experts who administer intelligence tests and those who rely on test scores when making decisions about people share similar opinions on the nature of intelligence. Snyderman and Rothman (1987) conducted a survey of over 750 experts (e.g., professors, psychologists, and educational specialists) and asked each participant whether he or she felt that there was a consensus among trained professionals as to the types of behaviors and characteristics that are designated "intelligent." While 53 percent agreed that there was a consensus, 39.5 percent felt that a consensus had not been achieved. The next question required respondents to review a list of 13 descriptors and to check all that they felt were important elements of intelligence. They were also asked to indicate whether or not current intelligence tests adequately measure each descriptor. Respondents were provided with space to write down any additional descriptors they felt had been omitted from the original list.

Results of the survey are presented in Table 44.1 and suggest that there is a good deal of consensus among experts concerning the key elements defining intelligence. However, characteristics such as adaptation to the environment and mental speed are clearly not measured adequately by intelligence tests that were in use at the time of the survey.

Any discussion of intelligence and its assessment would be incomplete without mentioning the heritability of general intelligence and its contemporary marker, the IQ score. When all is said and done, this controversial topic boils down to determining the relative contributions of nature (genetics) versus nurture (environmental influences) in explaining individual and group differences in "*g*" and in intelligence test performance. This is a topic of interest because it has been reported that in terms of average IQ, Jewish Americans perform best followed in order by Asian, white, Latino, and African Americans (Herrnstein & Murray, 1994). The discussion that follows draws heavily on the work of Brody (1992), Gottfredson (2005), Nisbett (2005), Reynolds (2000), and Rushton and Jensen (2005). Three basic positions that address the nature versus nurture controversy are presented:

1. *Culture-only theory:* Differences among whites, blacks, Hispanics, Asians, and other groups result exclusively from dissimilarities in environmental and cultural experiences that reflect inequalities across groups in terms of educational, nutritional, economic, and social opportunities. This hypothesis, which may be attributed in part to the work of the behavioral psychologist J. B. Watson, predicts that gaps between the groups on measures of "*g*," standardized intelligence scales, and academic achievement tests, to mention only a few of the possible indicators, will increase or decease depending on the degree of similarity or dissimilarity of environmental and cultural opportunities. Genetic heritage is, for the majority of people, irrelevant.

Table 44.1 Descriptors of human intelligence

Descriptor	*% checking important*	*% checking inadequately measured*
Abstract thinking or reasoning	99.3	19.9
Problem-solving ability	97.7	27.3
Capacity to acquire knowledge	96.0	42.2
Memory	80.5	12.7
Adaptation to environment	77.2	75.3
Mental speed	71.7	12.8
Linguistic competence	71.0	14.0
Mathematical competence	67.9	12.1
General knowledge	62.4	10.7
Creativity	59.6	88.3
Sensory acuity	24.4	57.7
Goal directiveness	24.0	64.1
Achievement motivation	18.9	71.7

SOURCE: Adapted from Snyderman and Rothman, 1987.

In the United States over the past 80 years, whites have consistently earned IQs that are about 15 points higher than those of blacks (mean IQs of 100 and 85, respectively). However, there is evidence that this disparity has declined in recent decades and may actually be eliminated within 20 years. One can only speculate as to why this change is occurring, but it supports the culture-only hypothesis. Advocates of this position attribute the finding to improved nutrition, education, and socioeconomic status among African Americans, along with reduced discrimination and prejudice toward blacks on behalf of the government and majority population. Examination of mean IQs from the WISC-IV, which was published in 2003, reveals a clear rank ordering of the racial groups in the standardization sample as follows: Asians (106.5), whites (103.2), Hispanics (93.1), and African Americans (91.7). These data indicate that the IQ gap between whites and blacks on this highly regarded measure of intelligence has narrowed by .23 of a standard deviation (Weiss, Saklofske, Prifitera, & Holdnack, 2006).

Additional support for the culture-only hypothesis is cited in Nisbett (2005) and involves a comparison of IQs from German children fathered by African American military personnel with German children fathered by white American solders. The groups of fathers demonstrated the expected white-black IQ disparity on tests of intelligence. However, the children of white GIs and black GIs had average IQs of 97 and 96.5, respectively. This finding suggests that the black-white IQ gap seen in American samples is not genetically determined.

2. *Hereditarian theory:* Differences across racial groups are due to an interaction of genetic and environmental influences that are roughly 50 percent genetic and 50 percent environmental. In part, this hypothesis may be attributed to Charles Darwin and Sir Frances Galton and predicts that average group differences in measured "*g*" and other ability tests will remain, albeit somewhat diminished, if all racial groups are treated equally and have identical environmental and cultural opportunities. Evidence supporting this position comes from adoption studies, investigations of monozygotic and dizygotic twin pairs, and studies demonstrating that "*g*" and IQ are rooted substantially in biology.

In one investigation, researchers obtained IQ scores on three groups of participants: biological fathers whose sons were adopted, stepfathers, and adult sons. The intelligence tests were administered during military induction of the fathers, stepfathers, and sons. The correlation between biological fathers and their adult sons was .20, whereas the association between IQ scores of adoptive fathers and their adult stepsons was a nonsignificant .02. This finding suggests that adopted children, even after reaching adulthood, are more likely to resemble their biological (genetic influences) fathers than their adoptive fathers (environmental influences).

Monozygotic twins develop from a single egg and have the same genetic makeup (identical twins). Conversely, dizygotic pairs originate from two eggs and have the genetic similarity of siblings (fraternal twins). If IQ and "*g*" are influenced by genetics, monozygotic twins should be more alike in terms of performance on ability measures than are dizygotic twins. This appears to be the case because the IQs of identical twins reared together correlate .86, whereas the correlation is .55 for fraternal twins reared together. For twin pairs reared apart, the IQ correlations are .76 and .35 for identical and fraternal twins, respectively. These patterns and levels of association, which are reported in Sattler (2001), suggest that, as far as IQ is concerned, genetic factors account for a major proportion of the variance.

There is a large body of research demonstrating that general intelligence and biological characteristics are significantly associated. In a recent meta-analysis involving 37 samples with a total of 1530 individuals, Mc Daniel (2005) reported that brain volume and intelligence as measured by standardized ability tests are correlated .41 for adult women and .38 for adult men. A reasonable conclusion from this line of research is that bigger brains are more efficient because they contain more neurons and synapses than smaller brains. There are at least two studies reviewed by Brody (1992) of significant associations (*rs* = .40 and .42) between IQ and conduction velocities in the median nerve of the arm. This measure of peripheral nervous system function is purely biological and involves no cognitive activity on the part of the study participants. Finally, inspection time (IT; time required to become aware of a stimulus by varying exposure time) in undergraduate students correlates –.74 with a popular test of fluid intelligence, suggesting that IT reflects a neural-processing aspect of intellectual functioning (Osmon & Jackson, 2002).

3. *Test bias theory:* This position assumes that intelligence scales and other ability tests (e.g., Scholastic Aptitude Test [SAT]) systematically underestimate the knowledge, aptitudes, and skills of minority members because they are poorly designed and culturally loaded in favor of middle-class persons of Euro-American descent. This position is a variant of the culture-only hypothesis and lacks empirical support. Nevertheless, it is so frequently asserted by critics of intelligence testing that it is worthy of independent discussion. Critics of psychometric assessment base their claims of test bias on mean score differences across racial, ethnic, and socioeconomic groups or the presence of individual items within a test that appear to be unfair or insulting to minority members.

The "mean score difference" definition of test bias has been widely researched and found to be without merit. Although groups of interest may differ in level of performance, studies of intelligence measures such as the Wechsler scales have repeatedly found little or no differences across ethnic/racial/sociocultural groups in terms of item order difficulty levels, internal consistency, and patterns of scoring determined by computation of the factorial structure of the test. Another approach to assessing bias in an IQ test is to evaluate its ability to predict academic achievement (or some other relevant variable such as job performance) for different ethnic, racial, and/or sociocultural groups. This is typically accomplished by first generating regression lines to predict academic achievement scores on the basis of IQ for each group and than comparing these lines statistically. Numerous investigations have demonstrated that widely used ability tests, including the Wechsler scales and Stanford-Binet, in most instances predict achievement scores in an effective and unbiased manor. However, an IQ test is said to be biased against a specific ethnic group if it systematically underpredicts academic achievement and performance for those individuals. If the test score in question is used to deny placement in a gifted program or admission to college, this would be an unacceptable situation since, if given the opportunity, these examinees are likely to perform better than expected based on the test scores alone. However, the only instances where bias in prediction has emerged involved overpredicting the achievement of minority persons. When researchers compared whites and minorities, the actual achievement of the latter individuals was often lower than predicted using a regression equation developed using both groups together.

The most common claim of test bias involves subjective examination of individual test items. This approach may identify potentially offensive content but it is ineffective for selecting biased items. Subjective examination is simply a matter of opinion as to whether an item is unfair, offensive, or biased for one or more groups in the population. A number of investigations have demonstrated that "expert" representatives from the majority and minority communities performed at no better than chance levels, and in some cases below chance levels, when they attempted to identify biased test items. A classic example of "expert" judgmental error involves an item taken from the Wechsler Intelligence Scales for Children. The child is asked what one should do if a much smaller child starts to fight with you. It was the opinion of some "experts" that this item was biased against African Americans because they felt that urban black children are taught by their parents and friends to strike smaller children who are showing aggression toward them. Statistical analysis of item difficulty levels for African American and Euro-American children who were asked this question indicated that the former group found the item easier than did the latter group. In a second study with an entirely new sample of participants, the "fight" item was passed by 73 percent of African American children and by 71 percent of Euro-American children (Sattler, 2001). Obviously, subjective inspection of test content is ineffective for identifying which items are more difficult for one ethnic group compared to another. However, review of test items by "experts" is worthwhile if it succeeds in eliminating insensitively worded or offensive items from future tests.

APPLICATIONS

There is little doubt that the construct of human intelligence and its assessment has had a major impact on societies as well as on individuals. For example, Lynn and Vanhanen (2002) proffered the view that IQ is one of the more important determinants of national economic development. They took the well-established fact that IQ and socioeconomic status are positively and significantly correlated for individuals and extended this idea to nations. After deriving IQ scores (from the average of various intelligence tests) for 81 countries, they reported highly significant associations between IQ and measures of per capita gross domestic product (GDP). Although the nations with the lowest IQs also had the lowest GDPs, the picture was more variable for high-IQ countries. Those with GDPs lower than expected on the basis of their IQs typically had a history of imposed economic regulation (e.g., recent or current Communist political structure). Information of this type may prove useful to world leaders who are constantly trying to identify and eliminate problems associated with globalization.

For individuals, the construct of intelligence and its measurement has a great deal of importance because ability tests are utilized in numerous settings such as schools, hospitals, clinics, and employment offices. Based on a survey of 277 clinical psychologists (Harrison, A. S. Kaufman, Hickman, & N. L. Kaufman, 1988), the foremost reason for administering an intelligence test was to estimate the person's potential or cognitive capacity. Three additional purposes for giving an intelligence test include (a) the need to obtain clinically relevant information about cognitive strengths and weaknesses, (b) assessing the functional integrity of the brain, and (c) assisting in determining appropriate vocational or educational placement. Today, adolescents and adults referred for cognitive evaluation will almost certainly be given the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III). Because the WAIS-III is one of the most thoroughly researched and frequently administered individual intelligence tests (Kaufman & Lichtenberger, 2006), a brief description of the scale is provided in Table 44.2.

Trained examiners must administer, score, and interpret the WAIS-III, with testing times ranging from approximately 60 to 120 minutes. Administration time varies depending on the overall level of ability of the examinee and whether a neurological or psychiatric disorder is part of the clinical picture. The WAIS-III is typically administered as part of a test battery. For example, if an examinee is referred for vocational assistance, tests of occupational interests and specific aptitudes (e.g., mechanical, dental, etc.) may supplement the WAIS-III. For examinees with a neurological disorder involving the brain (e.g., suspected Alzheimer's disease or history of significant head injury), measures of memory, constructional praxis, and executive functions are likely to be included in the test battery.

FUTURE TRENDS

In the future, although no universal definition of intelligence will be formulated, it is certain that basic research will continue unabated, as will standardized testing in educational, clinical, and employment settings. Interest in the biological and sensory-motor correlates of human ability, reminiscent of Galton and associates, will reemerge as a scientifically respectable line of research. As popular tests are revised and updated, there will be attempts to link research, theory (especially from cognitive psychology), and practice. Traditional IQs will be further deemphasized, but the old standards of general information, attention-concentration, vocabulary knowledge, abstract thinking, and various forms of nonverbal problem solving and processing speed will continue to be assessed. The WAIS-IV will follow the WISC-IV and drop both the Verbal and Performance IQs when it is published sometime between 2008 and 2010. Although the WAIS-IV Full Scale IQ will remain, it will be deemphasized and attention will be given to Indexes that measure separate capacities of intelligence within a theoretical framework. Future tests will likely include components or subtests that assess effective intelligence, both within and independent of dominant cultural influences. Finally, it is

Table 44.2 Description of the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III)

Verbal Subtests	*Requires*
Vocabulary:	Defining words
Similarities:	Explaining similarities between words or concepts
Arithmetic:	Mentally solving basic arithmetic problems
Digit Span:	Recalling series of digits presented orally
Information:	Answering general knowledge questions
Comprehension:	Answering questions requiring common sense
Letter-Number Sequencing:	Mentally rearranging and recalling number and letter series
Performance Subtests	*Requires*
Picture Completion:	Identifying missing parts in pictures
Digit Symbol-Coding:	Rapidly copying symbols that are paired with numbers
Block Design:	Reproducing abstract patterns using colored blocks
Matrix Reasoning:	Solving nonverbal reasoning tasks
Picture Arrangement:	Arranging series of pictures in logical sequences
Symbol Search:	Rapidly discriminating among nonsense symbols
Object Assembly:	Constructing jigsaw puzzles
Composites Indexes	*Measures*
Verbal Comprehension:	Comprehension and reasoning with verbal material
Perceptual Organization:	Reasoning with visual material
Working Memory:	Mentally manipulating orally presented material
Processing Speed:	Rapid processing of visual material
Composite IQs	
Verbal Scale:	Estimated overall verbal ability
Performance Scale:	Estimated overall nonverbal ability
Full Scale:	Estimate of overall ability or "g"

hoped that the major test publishers will include measures in their scales that tap variables not traditionally viewed as relevant. Lykken (2005) noted that extraordinary mental energy is a characteristic of great intellects (e.g., Galton, Picasso, Napoleon, and Einstein) and that it is an essential component of any formula for exceptional accomplishment. If a valid and reliable measure of mental energy were devised and included as part of standard test administration, the product of intelligence (e.g., Full Scale IQ) and mental energy might be a better predictor of academic and occupational success than IQ alone.

REFERENCES AND FURTHER READINGS

Borsboom, D., & Dolan, C. V. (2006). Why *g* is not an adaptation: A comment on Kanazawa (2004). *Psychological Review, 113,* 433–437.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Brody, N. (2003). Construct validation of the Sternberg Triarchic Abilities Test: Comment and analysis. *Intelligence, 31,* 319–329.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122–130). New York: Guilford.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gottfredson, L. S. (2002). *g*: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.

Gottfredson, L. S. (2003a). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence, 31,* 334–397.

Gottfredson, L. S. (2003b). On Sternberg's "Reply to Gottfredson." *Intelligence, 31,* 415–424.

Gottfredson, L. S. (2004). Life, death, and intelligence. *Journal of Cognitive Education and Psychology, 4,* 23–46.

Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law, 11,* 311–319.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13,* 1–4.

Harrison, P. L., Kaufman, A. S., Hickman, J. A., & Kaufman, N. L. (1988). A survey of tests used for adult assessment. *Journal of Psychoeducational Assessment, 6,* 188–198.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: The Free Press.

Kaplan, R. M., & Saccuzzo, D. P. (1997). *Psychological testing: Principles, applications, and issues* (4th ed.). New York: Brooks/Cole.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). New York: Wiley.

Lykken, D. T. (2005). Mental energy. *Intelligence, 33,* 331–335.

Lynn, R., & Vanhanen, T. (2002). *IQ and wealth of nations.* Westport, CT: Praeger.

McDaniel, M. A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence, 33,* 337–346.

Nisbett, R. E. (2005). Heredity, environment, and race differences in IQ: A commentary on Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11,* 302–319.

Osmon, D. C., & Jackson, R. (2002). Inspection time and IQ: Fluid or perceptual aspect of intelligence? *Intelligence, 30,* 119–127.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 Members. *Archives of Clinical Neuropsychology, 20,* 33–65.

Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law, 6,* 144–150.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11,* 235–294.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.

Sattler, J. M., & Ryan, J. J. (2001). Wechsler Adult Intelligence Scale-III (WAIS-III): Description. In J. M. Sattler, *Assessment of children: Cognitive applications* (4th ed., pp. 375–414). San Diego, CA: Author.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42,* 137–144.

Sokal, M. M. (1987). James McKenn Cattell and mental anthropometry: Nineteenth-century science and reform and the origins of psychological testing. In M. M. Sokal (Ed.), *Psychological testing and American society: 1890–1930* (pp. 21–45). New Brunswick, NJ: Rutgers University Press.

Spearman, C. E. (1923). *The nature of intelligence and the principles of cognition.* London: McMillan.

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1997). *Successful intelligence.* New York: Plume.

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond *g*: Putting multiple intelligences theory to the test. *Intelligence, 34,* 487–502.

Watson, R. I. (1963). *The great psychologists.* Philadelphia: Lippincott Company.

Wechsler, D. (1939). *The measurement of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-Third Edition.* San Antonio, TX: The Psychological Corporation.

Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WISC-IV: Advanced clinical interpretation.* New York: Elsevier.

45

Motivation and Emotion

Joseph J. Palladino and Christopher M. Bloom

University of Southern Indiana

The two topics that comprise this chapter have long histories in psychology. One way to assess interest in these topics is to look at the place they have held and currently hold in the undergraduate psychology curriculum. On one hand, course offerings on motivation seem to be declining; teachers and researchers have incorporated some topics subsumed under the word *motivation* within the increasing number of courses in behavioral neuroscience. On the other hand, there appears to be growing interest in the topic of *emotion*, with courses on the topic being offered as well as a number of special topics-related courses such as the psychology of fear. Certainly, the relatively new (historically speaking) "positive psychology" movement has focused more attention on some of the topics that would be listed as the province of emotion.

MOTIVATION

One common definition of motivation is the "physiological and psychological factors that account for the arousal (energizing), direction, and persistence of behavior" (Davis & Palladino, 2007, p. 231). The word motivation is derived from the Latin word meaning "to move." In essence, motivation addresses the causes of behaviors; the key questions focus on what "moves" organisms to act in the ways that they do. Most attempts to explain differences in level of motivation fall into a few general theoretical perspectives.

Theoretical Perspectives on Motivation

One of the oldest perspectives, *drive theory*, proposes the existence of internal motivational states called *drives* that result from some physiological need. Consequently, drive theory is most relevant to the biologically related motives such as hunger and thirst. Drive-reduction theory proposes that drives direct our behavior to reduce an existing state of physiological imbalance (which results from a state of deprivation) and return us to a more balanced state of *homeostasis*. From a behavioral standpoint, behaviors that lead to drive reduction are therefore reinforced and are more likely to occur again.

Optimum-level theories feature an opposite approach and propose that organisms are motivated not by drive reduction but by a desire to maintain some optimum level of arousal. This optimum level of arousal varies from individual to individual, thus accounting for the variability we see in human and animal behavior. One classic example of this approach is seen in research on sensation seeking, which is used to account for a wide range of behaviors: food preferences, gambling, drug and alcohol use, sexual behavior, occupations, and interests in sky diving and race car driving (Flory, Lynam, Milich, Leukefeld, & Clayton, 2004; McDaniel & Zuckerman, 2003; Zuckerman, 1994).

Perspectives on motivation also include *cognitive consistency theories*, which propose that we are motivated by a desire to maintain a state in which our thoughts and actions are consistent, a form of what might be termed psychological homeostasis. When thoughts and behaviors are inconsistent we experience *cognitive dissonance*, and we are motivated to reduce it by altering our thinking about the behaviors that are causing the cognitive dissonance. For example, a person who smokes may find this behavior incompatible with information on the dangers of smoking, thus creating cognitive dissonance. As a result, this smoker might focus on people who have smoked for long periods of time without ill effects in an effort to reduce the dissonance.

Another approach to motivation suggests that characteristics of the goals or incentives in our environment determine behavior. Thus, rather than being driven by internal forces (drive theory) we are *pulled* by external forces. Over many decades, a great deal of attention has focused on Abraham Maslow's hierarchy of needs, which proposes the existence of the following hierarchy: basic physiological needs (e.g., for food, water, and sleep), safety needs, belongingness and love needs, esteem needs, and self-actualization. According to Maslow, we must first satisfy lower needs (especially basic physiological needs) and then strive to satisfy successively higher needs; consequently, his theory is sometimes described as a growth theory of motivation.

Hunger

Of all of the specific motives that we might discuss, hunger has received considerable attention in the academic and popular literature. In fact, in recent decades, we have seen growing concern over the accelerating rates of obesity with its concomitant negative effects on health and well-being.

One of the oldest theories of hunger focused on the role of the tiny hypothalamus in the brain, which was thought to contain a "start-eating" center along with a "stop-eating," or satiety, center. Special receptors in the hypothalamus were thought to be sensitive to circulating levels of glucose in the blood. Surgical lesions of two separate parts of the hypothalamus tended to yield two different results: one lesion led to reduced food intake, a second lesion led to increased food intake. Another early view of the cause of hunger was that it was related to the amount of stored fat. When levels of fat were drawn down, signals would be sent to the brain to initiate eating in order to restore the fat that had been used. Subsequent research suggests that these approaches to understanding hunger are oversimplified. We now know that hunger has multiple biological determinants, including the hormone leptin as well as hormones in the small intestine that seem to serve as a stop-eating signal (Holtkamp et al., 2003; Nash, 2002).

Much of the current interest in hunger focuses on overeating (although anorexia nervosa and bulimia nervosa are also of great concern), which is a much more serious worldwide problem in most developed countries than is starvation. According to the U.S. Department of Health and Human Services (2001), obesity is a significant risk factor for coronary heart disease, high cholesterol, stroke, hypertension (high blood pressure), osteoarthritis, sleep disorders (especially sleep apnea), and Type II diabetes. What's more, obesity is associated with increased risk for the development of several types of cancer including colon, gallbladder, prostate, and kidney (U.S. Department of Health and Human Services, 2001; World Health Organization, 2005). The impact of obesity on health can be seen clearly in the observation that obesity is implicated in 4 of the top 10 causes of death in the United States (National Center for Health Statistics, 2004).

Obesity is a worldwide problem; the World Health Organization estimates the number of obese and overweight people in the world is more than one billion, with at least 300 million of these people considered to be obese. Worldwide there are an estimated 22 million children under age five who are overweight. Obesity rates vary quite a bit around the world. These rates tend to be quite low in China and Japan and quite high in urban Samoa (World Health Organization, 2005), which suggests that cultural and social factors are involved in what we weigh. For example, rates of obesity are higher among those in the lower socioeconomic classes than among those in the middle and upper socioeconomic classes (Bray, 2004).

Body Mass Index

But just how do we know if we are overweight or obese? The original method used to make such judgments involved use of tables of desirable weight (listed by height) published by major insurance companies. Unfortunately, these tables were based on the weights of people who could afford to buy the insurance policies, thus the sample was hardly representative of the entire population (most likely they were in better health than the general population). What's more, as the population added pounds across the decades, additional weight was added to the tables of desirable weight, perhaps providing some small comfort as we added inches to our girth!

A major change in the way we assess our weight occurred in 1998 when the National Heart, Lung, and Blood Institute (NHLBI) devised a new approach called *body mass index* (BMI). The BMI index is a single number (most people have a BMI in the 20 to 40 range) that represents your height and weight without reference to sex. You can determine your BMI by entering your height and your weight into one of many Web sites that are available to make such calculations. For example, see http://www.nhlbisupport.com/bmi.

But why would you want to know your BMI? Simply put, your BMI is a better indicator of health and disease risk than other weight-related measures. For example, an increase of just 1 in the BMI increases heart disease risk by 5 percent in men and 7 percent in women (Kenchaiah et al., 2002). Nevertheless, the BMI should not be used by everyone. For example, competitive athletes and body builders should not use the BMI because their high muscle mass will give a misleading picture of health-related risk. Why? Muscle tissue weighs more than fat; hence, the picture of health for a bodybuilder based on BMI can be misleading. The BMI should also not be used by pregnant women and those who are frail and sedentary. Although the exact points on the BMI scale that trigger recommendations to lose weight are still debated (Strawbridge, Wallhagen, & Sheman, 2000), the existing NHLBI guidelines are as follows:

Underweight: less than 18.5
Normal weight: 18.5–24.9
Overweight: 25–29.9
Obese: BMI of 30 or greater

The most recent surveys of the prevalence of overweight and obesity in the U.S. were completed in 2003–2004. The percentage of adults age 20 years and over who were overweight or obese (BMI of 25 or greater) was 66 percent. The percentage of adults who were obese (BMI over 30) was 32 percent (National Center for Health Statistics, 2007).

Causes of Obesity

Although the original focus on hunger was on the role of the hypothalamus in gauging glucose levels, it is possible that tumors in the hypothalamus and elsewhere in the endocrine and nervous systems might be responsible for cases of obesity. The evidence indicates, however, that only a small percentage of cases of obesity (about 5 percent) have various endocrine problems or damage to the hypothalamus (Bray, 2004). On the other hand, evidence for the role of genetic factors in weight and obesity is quite strong (Bulike, Sullivan, & Kendler, 2003, Schouseboe et al., 2004). For example, there is a high correlation between the weights of identical twins, even when they are separated at birth. Such findings point strongly to a genetic influence on what we weigh. But what exactly is it that we might inherit? One possibility is that genetic factors can influence what is termed the *basal metabolic rate* (BMR). Think of the BMR as the pilot light on a stove or a water heater—both devices will burn fuel for energy even when no hot water is being used or the stove is not turned on and in use. Likewise, throughout the day our body needs energy to keep us alive—heart rate, breathing, and so on. The rate at which we burn energy in the body to keep basic processes going is our BMR. Several factors influence our basal metabolic rate: Men have a higher rate of BMR than women, and the BMR drops with age. This latter finding explains why we tend to put on weight as we age, even if we maintain our same level of exercise (Goodenough, McGuire, & Wallace, 2005). As we age the body uses fewer calories to keep us going; any excess calories will have a tendency to go directly to our hips and abdomen!

One way to look at the problem of obesity is to view it as simply the consequence of taking in more calories than we burn. In general, it takes 3,500 calories to equal a pound of fat (Goodenough et al., 2005). Thus, if we consumed 10 more calories than we burn every day for a year we would add 10 pounds. Multiple this result by several years and it is easy to see how we can add a significant amount of weight over time as we age.

Dieting

Listen to television, read your e-mail, check your snail mail. The odds are you will come across information on a new weight-loss diet. According to "New Diet Winners" (2007), "Americans don't give up easily. Those hoping to lose weight have put a whole new crop of diet books on the best-seller list" (p. 12). Despite the multiple determinants of obesity, the number one plan for losing weight is a diet of some kind (Kaplan, 2007). *Consumer Reports* recently published an analysis of a number of diet books, with mixed views on their effectiveness.

One problem with many diets over the years is they may be highly unrealistic as long-term solutions to obesity. They may work initially (often the result of calorie reduction); however, in the end they are likely to fail, especially if they require that you consume foods that are not very desirable. How long could anyone follow a hot dog diet (just hot dogs and condiments), a pumpkin and carrot diet, a mushroom diet, a milk-only diet, or an eggs and wine diet? The list of highly unusual diets goes on and grows longer every year, yet our insatiable desire to lose weight finds a ready audience. In fact, research on the effectiveness of dieting has yielded two general conclusions: (a) Diets work in the short term, and (b) Diets do not work in the long run (Mann et al., 2007). What's more, "The more time that elapses between the end of a diet and the follow-up, the more weight is regained" (Mann et al., 2007, p. 211).

When we diet, we are engaged in a battle with the consequences of evolution. Over thousands of years, evolution has favored those who ate as many calories as possible to protect against the next unpredictable famine. Thus, they were in a better position to pass on their genes to the next generation. Evolution has led us to eat when food is available, and in many countries around the world food is very readily available. Consider the following: Between 1984 and 2000, our daily intake of calories increased by 500 per day. We tend to eat out more often, consuming more fat, less fiber, more cholesterol, and more calories than are found in homemade meals (Spake, 2002). What's more, we live in an era of supersizing. In fact, just increasing the portion size will lead us to consume more food (Rolls, Roe, Kral, Meengs, & Wall, 2004).

But are there some general principles that can help us in the battle of the bulge? As "New Diet Winners" (2007) notes, "The basic formula for losing weight has not changed: Consume fewer calories than you burn—about 500 fewer every day, to lose about a pound a week" (p. 12).

Consider the following if you intend to look for ways to lose weight for the long haul:

1. *Increase physical activity.* A club membership at the local gym is not required, although the surge in memberships after New Year's Day attests to their appeal at one time of the year. Consider the simplest exercise possible—walking. Buy a pedometer and set a goal (10,000 steps per day is a reasonable goal). Build physical exercise into your everyday routine, so you do not have to make a special trip to exercise. Use the stairs, not the elevator. Park some distance from your destination at work or the mall, thus forcing yourself to walk a few hundred extra steps each way. Across an entire year, a few hundred extra steps a day can add up to a weight loss of several pounds.
2. *Eat slowly.* (This does not mean chewing your food 27 times before you swallow!) The additional time may allow the body's biological signals to indicate that you are really full. One added benefit of slowing down your eating is you might find out how the food actually tastes!
3. *Identify cues that lead you to eat.* Some of us eat when we are upset, or when we watch television. Be aware and make changes. Be careful to avoid the influence of cues. We are highly visual animals and readily respond to the sight of food (note how good everything looks when we are in a supermarket—even the frozen food that never measures up to the picture on the package).
4. *Diets often remove the enjoyment from eating.* Add some spice (literally) to what you eat in order to make any diet more effective in helping you lose weight.
5. *Try to preload your stomach with bulky, low-calorie foods* such as popcorn (no butter), carrots, or celery.
6. *Do not go grocery shopping when you are hungry*, otherwise you are likely to come home with a basketful of items that you do not need and should not eat.

Other Eating Disorders

Although a great deal of recent attention has focused on overeating and the health effects of obesity, there are other eating-related disorders that should be considered. Mark Chavez and Thomas Insel (2007) of the National Institute of Mental Health recently wrote, "Advancing the understanding and treatment of eating disorders is an issue of immense public health importance and is recognized as an area of high priority by the National Institute of Mental Health (NIMH)" (p. 159). The primary eating-related problems beyond obesity are *anorexia nervosa*, *bulimia nervosa*, and a category of *eating disorder not otherwise specified* (American Psychiatric Association, 2000).

Anorexia nervosa literally means "nervous loss of appetite"; however, this literal meaning is misleading because victims of anorexia nervosa deny their hunger as result of an intense fear of becoming fat (Keel, 2005). Anorexia nervosa tends to occur in women, especially between the ages of 13 and 20 (American Psychiatric Association, 2000). It can be diagnosed when a person's weight falls below 85 percent of a desirable weight, or has a BMI of less than 17.5 (Wilson, Grilo, & Vitousek, 2007). Most victims, however, tend to lose far more weight than these figures would suggest. In fact, it is not uncommon for victims to take on the appearance of a skeleton, wrapped in tissue-like skin with virtually every bone visible; the loss of weight is so severe that it may be difficult to discern the person's sex when looking at the naked body. This "semi-starvation brings about profound and predictable changes in mood, behavior, and physiology. These include depression, social withdrawal, food preoccupation, altered hormone secretion, amenorrhea, and decreased metabolic rate" (Wilson, Grilo, & Vitousek, 2007, p. 199). There is a long list of associated medical problems, including anemia, low blood pressure, and stress fractures, that result when the calcium is extracted from bones (American Psychiatric Association, 2007). Moreover, changes in potassium and sodium levels in the body can lead to cardiac arrest or kidney failure. Thus, it is not surprising that the mortality rate from anorexia nervosa is approximately 5 percent per decade (Birmingham, Su, Hlynsky, Goldner, & Gao, 2005), which is the highest mortality rate of all mental disorders (Millar et al., 2005).

People suffering from *bulimia nervosa* (literally "continuous nervous hunger") may consume a tremendous number of calories at one sitting and then vomit, take laxatives or diuretics, or engage in excessive exercise (American Psychiatric Association, 2000). These behaviors following an episode of *binge eating* are referred to as *compensatory behaviors* and are the mark of the *purging* component of bulimia nervosa. There are also bulimics who restrict their food intake and exercise for weight control. "Associated general psychopathology (e.g., depression and personality disorders) and psychosocial impairment are common" (Wilson et al., 2007, p. 203). The disorder tends to occur in young women. Despite a long list of medical complications that include irregular heartbeat and altered levels of potassium, sodium, and calcium, it is quite often a dentist who first notices the disorder. In contrast to persons with anorexia nervosa, most victims tend to maintain a body weight very close to their desirable weight; however, the stomach acid that accompanies frequent vomiting is likely to lead to an increased number of cavities.

EMOTION

It was in 1884 that William James first posed the question, "What is an emotion?" and 125 years later psychologists and neuroscientists continue to seek an answer. In fact, more than 90 definitions of the term were offered during the 20th century (Plutchik, 2001). There is general agreement that the word *emotion* is derived from the Latin word meaning "to move" and from the prefix *e*, which means "away." This derivation from the Latin word for "move" reveals why the topics of emotion and motivation have been linked historically in textbooks and in other writings.

The question of what is emotion, like emotion itself, is multifaceted, spawning centuries of research from multiple theoretical perspectives. These perspectives differ not only in school of thought but the aspect of emotion on which they focus. When we talk about our own emotions, we typically are really describing how we feel. (Note that the term *mood* is often used when the feelings associated with emotions last for an extended period of time.) Indeed, the subjective internal experience is an important aspect of emotion, but it is not the only facet that interests researchers. Bodily changes, cognitions, and the way we communicate via emotional expression are all important areas of inquiry in the study of emotion. Whatever emotion is, it is an integral part of our lives. Oatley and Duncan (1994) asked participants to keep a log of their emotional experiences. Participants reported having at least one emotional experience per day that was strong enough to elicit a noticeable physical reaction. One third of the participants reported that this emotional experience lasted 30 minutes or more! No wonder emotion has garnered so much attention from novelists, poets, filmmakers, and psychologists.

Evolution and Emotion

The importance of emotion was not lost on Charles Darwin, who focused on the expressive aspects of emotion. He believed, not surprisingly, that emotion could only be understood as part of the evolutionary history and survival of a species. His focus was not, however, on the role of emotional expression in communication between members of a species. Instead, he believed that emotional expression served another purpose in survival. Darwin's "principle of serviceable habits" stated that facial expressions had directly serviced the organism. He used the facial expression we associate with disgust as an example to make his point. Tonight, when you brush your teeth, take a look at the expression you make as you spit out your toothpaste. Darwin noted that it is the same face we make when the emotion of disgust is elicited. According to Darwin, this expression originally served the purpose of spitting out spoiled or poisonous foods and through habit came to be produced whenever a disgusting notion comes to mind. In other words, over time disgust probably came under some degree of voluntary control; consequently, the facial expression of disgust now occurs as a response to circumstances that have less to do with the taste of food and more to do with our reactions to moral offenses, foul languages, or poor hygiene (Rozin, Lowery, & Ebert, 1994). Both anger and fear work similarly in Darwin's estimation. For instance, an angry cat arches its back to look larger, hisses to show its teeth, and lays its ears back to protect them from injury in a fight. These outward expressions of anger serve a real purpose, aside from their value in communication (Darwin, 1872/1965). Put another way, emotions can increase the chances of survival by providing us with a readiness for actions such as fighting predators that have helped people survive throughout our evolutionary history (Plutchik, 1993, 2001). Anger "intimidates, influences others to do something you wish them to do, energizes an individual for attack or defense and spaces the participants in a conflict" (Plutchik, 2001, p. 348).

Darwin (1872/1965) believed that the value of emotional expression was limited, however, to the "negative" emotions. Positive emotional expression such as smiling with joy served no purpose, in Darwin's view. Instead, as posited in his "principle of antithesis," positive emotional expression is a result of the absence of the negative emotion. When a cat is docile, its mouth is closed, its ears and tail are in an upward position, and the hair on its back is relaxed. Darwin saw this seemingly opposite bodily position as not being an expression of its own but instead simply the absence of the negative, survival-valuable behavior.

Darwin (1872/1965) saw these emotional expressions as a function of evolution and therefore as biological, not culturally learned. Using a survey he distributed to missionaries, he collected evidence that these expressions were the same across multiple cultures even where contact with the West was extremely limited. Contemporary researchers have done much to bolster Darwin's claim. Ekman and Friesen (1971) investigated emotional expression in the preliterate Fore tribe in New Guinea. Using photographs, the highly isolated Fore tribe members were able to successfully select the appropriate facial expression to a story read to them by the researchers. These people could not have learned the meanings of facial expressions through the media, as they had no access to any media at all.

Using a version of the game "name that emotion," Ekman (2003) found that people from Western and Eastern literate cultures, including Japan and the United States, associate similar emotions with certain facial expressions. Not surprisingly, a recent extension of this research with participants in the United States, India, and Japan found that people tend to recognize emotions more easily when they are viewing people from their own culture (Elfenbein, Mandal, Ambady, Harizuka, & Kumar, 2002), but even so, these findings support the notion that humans have innate, genetically wired templates for expressing different emotions. It makes sense that there should be similarities across cultures. Human beings belong to the same species: our brains, our bodies, our autonomic nervous systems, our hormones, and our sense organs are similarly constructed. Modern-day researchers seem to have arrived at the conclusion that six basic emotions are generally recognized across the world: anger, disgust, fear, happiness, sadness, and surprise, with agreement highest for facial expressions of happiness. Researchers have even seen emotional expressions on the faces of infants born blind and deaf, which clearly suggests that there is a strong biological/evolutionary component to what we call emotion. As Darwin suggested, emotional expressions appear to be universal.

The Emotional Body

Whereas Darwin focused on emotional expression, William James was captivated by the physiological component of emotion. James (1884) believed that the body is central to the generation of emotion. James sought to argue against the prevailing view that the experience of emotion generated bodily changes. Indeed, he believed the opposite to be true, that the bodily change created the emotional experience.

> The hypothesis here to be defended says…that the more rational statement is that we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be. (p. 190)

While James was crafting his theory, Danish physiologist Carl Lange was independently developing a similar notion of his own, resulting in the theory coming to be known as James-Lange theory. Lange also believed that bodily changes were the cause of emotion, but he took his theory a step further. Lange sought to determine which bodily manifestations accompanied each emotion. The idea that each emotion has its own physiological signature has come to be known as *autonomic specificity*. The term refers to Lange's focus on bodily changes controlled by the autonomic nervous system, such as blood pressure, respiration, and heart rate. Lange failed to provide compelling evidence to support this notion, but the idea has spawned countless investigations into the physiology of emotion. As physiological measurement has advanced, so has our understanding of the relation between bodily change and emotion. However, the James-Lange theory was not without its critics. One of the most vocal opponents of the theory was the physiologist Walter Cannon. Cannon argued that the type of body changes that James-Lange believed to be at the heart of emotion were nonspecific and too slow to be responsible for producing emotion. Cannon and Phillip Bard developed a competing theory that posited that the physiological change and the experience of emotion occurred simultaneously. Known as Cannon-Bard theory, it argued that physiological change did not precede emotion; instead, they emerged together to produce the experience we call emotion.

Thinking About Emotion

The cognitive aspects of emotion have been the focus of researchers as well. Cognitive theories of emotion have focused on the role of judgment and thought in the development of the emotional experience. Schacter and Singer (1962) proposed a theory that elaborated on William James's notions that bodily change equaled emotion. Titled the "two-factor theory of emotion," it didn't deny the importance of bodily change in the formation of emotion. Indeed, autonomic arousal was the "first" of the two factors. The theory differed from James's, however, in that it emphasized the importance of environmentally appropriate cognition as a key aspect of emotion. Once a bodily change is produced, the individual looks to evaluate the environment to explain the bodily state. That evaluation then results in the state being interpreted as a particular emotion. When a dog comes around the corner and surprises you as you walk down the street, the startle produces bodily change. Are you scared? Are you happy? The two-factor theory suggests that the answer depends on how you evaluate the dog. If you've had bad experiences in the past with dogs you may experience the bodily change as fear. A dog lover is more apt to experience the same physiological experience as happiness. Many modern theorists have also focused on the cognitive evaluation of stimuli as being vital in the experience of emotion. Lazarus (1991) proposed his cognitive-motivational-relational theory. The theory centers on the notion of appraisal of our environments. For Lazarus, appraisals focus on whether the environment has positive or negative personal implications. Further, the appraisals have two components, a primary and secondary appraisal. The primary appraisal determines whether an environmental stimulus has any personal relevance. Seeing a house cat walking down an alley is likely to be appraised as having little personal impact. Seeing a tiger, however, is likely to be appraised quite differently. Having appraised the tiger as quite relevant, the individual then begins the secondary appraisal process. The secondary appraisal focuses on how to maximize award or minimize harm in response to the environmental stimulus. In the case of the tiger, the individual's secondary appraisal would focus on how to escape the large predator unharmed.

Anatomy of Emotion

Since James, Lange, and Cannon emphasized the importance of physiology in the process of emotion, much has been learned about the neuroanatomical systems involved in affective experience. The human nervous system has two main parts, the central nervous system (CNS) and the periperhal nervous system (PNS). Each of these branches of the nervous system has a vital role to play in emotion.

The PNS is defined as all nervous system structures outside of the brain and spinal cord. It consists of a system of axonal fibers that bring sensory stimuli to the CNS (sensory neurons) as well as a response to said stimuli from the CNS to the rest of the body (motor neurons).

Theories of how emotions are generated abound but one aspect of emotional experience is undeniable: It is a biological organism that experiences the emotion. As such, there can be little doubt that an understanding of the anatomical structures involved in emotion is key to understanding the emotional experience in its totality.

A great deal of the focus on the nervous system and emotion involves, not surprisingly, the limbic system, which includes the amygdala, hippocampus, and the hypothalamus. The amygdala is a small almond-shaped structure that seems to react instantly to sensory inputs and can trigger the fight-or-flight response while the cortex is still evaluating the inputs. In a series of interesting experiments, Joseph LeDoux (1996) found that there are two pathways information can take to the brain. One route involves the cortex and the other is subcortical. But why would there be two routes to follow? As failing to respond to danger is more costly than responding inappropriately to a benign stimulus. The subcortical route is simply faster. In a related vein, researchers have found that people with a damaged amygdala rated faces with negative expressions to be as trustworthy as faces with positive expressions. What's more, brain scans reveal activation of the amygdala when people view facial expressions of fear and anger. The range of emotions for which the amygdala plays a role is expanding at a rapid rate; for example, research now indicates that the amygdala is activated while viewing pictures that portray sadness.

Emotion in Action

In more recent years, psychologists have developed the concept of *emotional intelligence*, which focuses on four qualities: (a) the ability to perceive emotions in others, (b) the ability to facilitate thought, (c) understanding emotions, (d) and managing emotions. The concept recognizes the fact that brainpower is not necessarily important for success and that the listed qualities have generally gone unrecognized. What's more, there is little to no relation between emotional intelligence and intelligence as measured by standard tests of intelligence. Researchers have found that scores of measures of emotional intelligence correlate with happiness when both people in a couple had high scores; happiness was in the mid-range if one member of the couple had a high score and the other had a low score.

REFERENCES AND FURTHER READINGS

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Rev. ed.). Washington, DC: Author.

Birmingham, C. L., Su, J., Hlynsky, J. A., Goldner, E. M., & Gao, M. (2005). The mortality rate from anorexia nervosa. *International Journal of Eating Disorders, 38*(2), 143–146.

Bray, G. A. (2004). Classification and evaluation of the overweight patient. In G. Bray & C. Bouchard (Eds.), *Handbook of obesity: Clinical applications* (2nd ed., pp. 1–32). New York: Marcel Dekker.

Bulike, C. M., Sullivan, P. F., & Kendler, K. S. (2003). Genetic and environmental contributions to obesity and binge eating. *International Journal of Eating Disorders, 33*(3), 293–298.

Chavez, M., & Insel, T. R. (2007). Eating disorders: National Institute of Mental Health's perspective. *American Psychologist, 62*, 159–166.

Darwin, C. (1965). *The expression of the emotions in man and animals.* Chicago: University of Chicago Press. (Original work published 1872)

Davis, S. F., & Palladino, J. J. (2007). *Psychology* (5th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Times Books.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124–129.

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-cultural patterns in emotion recognition: A meta-analysis. *Emotion, 2*, 75–84.

Flory, K., Lynam, D., Milich, R., Leukefeld, C., & Clayton, R. (2004). Early adolescent through young adult alcohol and marijuana use trajectories: Early predictors, young adult outcome, and predictive utility. *Development and Psychopathology, 16*, 193–213.

Goodenough, J., McGuire, B., & Wallace, R. A. (2005). *Biology of humans: Concepts, applications, and issues*. Upper Saddle River, NJ: Pearson Prentice Hall.

Holtkamp, K., Hebeband, J., Mika, C., Grzella, I., Heer, M., Heussen, N., et al. (2003). The effects of therapeutically induced weight gain in plasma leptin levels in patients with anorexia nervosa. *Journal of Psychiatric Research, 17*, 165–169.

James, W. (1884). What is an emotion? *Mind, 9*, 188–205.

Kaplan, R. M. (2007) Should Medicare reimburse providers for weight loss interventions? *American Psychologist, 62*, 217–219.

Keel, P. K. (2005). *Eating disorders.* Upper Saddle River, NJ: Pearson Prentice Hall.

Kenchaiah, A., Evans, J. C., Levy, D., Wilson, P. W. F., Benjamin, E. J., Larson, M. G., et al. (2002). Obesity and the risk of heart failure. *New England Journal of Medicine, 347*(50), 305–313.

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist, 46*, 819–834.

Mann, T., Tomiyama, A. J., Westling, E., Lew, A.-M., Samuels, B., & Chatman, J. (2007). Medicare's search for effective obesity treatments: Diets are not the answer? *American Psychologist, 62*, 320–322.

McDaniel, S. R., & Zuckerman, M. (2003). The relationship of impulsive sensation seeking in gambling activities. *Personality and Individual Differences, 35*, 1385–1400.

Millar, H. R., Wardell, F., Vyvyan, J. P., Naji, S. A., Prescott, G. J., & Eagles, J. M. (2005). Anorexia nervosa mortality in Northeast Scotland, 1965–1999. *American Journal of Psychiatry, 162*, 753–757.

Nash, J. M. (2002, September). Cracking the fat riddle. *Time,* pp. 46–55.

National Center for Health Statistics. (2004). *Health, United States, 2004*: *With chartbook on trends in the health of Americans.* Hyattsville, MD: U.S. Government Printing Office.

National Center for Health Statistics. (2007). *Prevalence of overweight and obesity among adults: United States, 2003–2004*. Retrieved on May 12, 2007, from http://www.cdc.gov/nchs/products/pubs/pubd/hestats/overweight/overwght_adult_03.htm

New diet winners. (2007, June). *Consumer Reports*, 12–17.

Oatley, K., & Duncan, E. (1994). The experience of emotion in everyday life: Brief report. *Cognition and Emotion, 8*, 369–381.

Plutchik, R. (2001). The nature of emotions. *American Scientist, 89,* 344–350.

Rolls, B. J., Roe, L. S., Kral, T. V. E., Meengs, J. S., & Wall, D. E. (2004). Increasing the portion size of a packaged snack increases energy intake in men and women. *Appetite, 42*(1), 63–69.

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology, 66*, 870–881.

Schacter, A., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review, 69*, 379–399.

Schousboe, K., Visscher, P. M., Erbas, B., Kyvik, K. O., Hopper, J. L., Henriksen, J. E., et al. (2004). Twin study of genetic and environmental influences on adult body size, shape, and composition. *International Journal of Obesity and Related Metabolic Disorders, 28*(10), 39–48.

Spake, A. (2002, August 19). America's "supersize" diet is fatter and sweeter—and deadlier. *U.S. News & World Report*, pp. 41–47.

Strawbridge, W. J., Wallhagen, M. I., & Sheman, S. J. (2000). New NHBLI clinical guidelines for obesity and overweight: Will they promote health? *American Journal of Public Health, 90*, 340–343.

Sullivan, P. F. (1995). Mortality in anorexia nervosa. *American Journal of Psychiatry, 152*, 1073–1074.

U.S. Department of Health and Human Services. (2001). *The Surgeon General's call to action to prevent and decrease overweight and obesity.* Rockville, MD: Author.

Wilson, G. T., Grilo, C. M., & Vitousek, K. M. (2007). Psychological treatment of eating disorders. *American Psychologist, 62*, 199–216.

World Health Organization. (2005). *Obesity and overweight*. Retrieved April 15, 2007, from http://who.int/dietphysicalactivity/publications/facs/obesity/en/print/html

Zuckerman, M. (1994). *Behavioral expression and biosocial bases of sensation seeking.* Cambridge, UK: Cambridge University Press.

PART VIII

Cognitive Psychology

46

MEMORY

A Look Into the Past, Present, and Future

JENNIFER M. BONDS-RAACKE AND JOHN RAACKE
University of North Carolina at Pembroke

Memory is, therefore, neither Perception nor Conception, but a state or affection of one of these, conditioned by lapse of time. —*Aristotle*, 350 BCE

Over two millennia have passed since Aristotle first began to reason about the properties and structure of memory. In addition to Aristotle, many other historical figures have spent time trying to decipher the complexities of memory. Plato was known to describe memory as a wax tablet to be written on. Additionally, philosophers such as Descartes and Kant portrayed simple and complex ideas as being the building blocks of thought and memory. Nevertheless, none of these great thinkers were actually capable of experimentally testing their beliefs. Only within the last century or so has the idea of memory come to be tested using scientific procedures. However, even with new advances in methodologies, at the heart of all memory research is the question that Aristotle attempted to answer many years ago: What is memory?

On the surface, this appears to be an easily answered question. Most people intuitively can provide an answer to this question. The most common response to this question when posed to introduction to psychology students is that memory is a storage space housed in the brain for remembering information. This answer is simple and noncomplex, and most people on the surface would agree that it does a pretty decent job of describing memory on some level. This becomes especially true when considering the universal definition that memory is the mental process of acquiring and retaining information for later recall. However, when one begins to ask simple questions about memory to help further define the concept, inadequacies in the aforementioned definitions become striking. For example, how does memory form? How do people store their memories? Where exactly is memory information stored? How much memory information can one store? How long can memory information be stored? How does one recall memories? Thus, although memory appears simple enough on the surface, it is far more complex than most realize.

THEORY AND METHODS

Researchers have proposed many theories (and accompanying methodologies) to explain how information is encoded, stored, and retrieved from memory. First, individuals who can be considered pioneers in memory research will be presented. Their contributions and current status in the field will be examined. Next, sensory memory will be covered, explaining how environmental stimuli are entered and retained in memory. Atkinson and Shiffrin's theory of short-term memory and Baddeley's working memory will be outlined and current empirical data regarding each presented. Finally, long-term memory will be examined including (a) the taxonomy of long-term memory, (b) flashbulb memories, and (c) autobiographical memories.

Pioneers in Memory Research

Herman von Ebbinghaus in the late 1800s was the first to take a scientific, systematic approach to the study of human memory. What was even more impressive than his being the first to undertake researching this topic was the ingenuity he demonstrated in developing his methodology and statistical procedures. Unlike many of today's researchers studying human memory, Ebbinghaus did not have a group of research participants. He also did not have a laboratory or technical research equipment. Rather, Ebbinghaus served as the sole participant on his studies of memory. One of his many inventions was the use of the meaningless consonant-vowel-consonant (CVC) triads. Ebbinghaus would learn a list of CVCs to a set criterion (e.g., one or two recitations without error), put aside the list for a period of time, and later relearn the list to the same criterion. He would then compute a savings score, which depicted the number of trials necessary for relearning. Ebbinghaus is famous for the graphical depiction of this information in the forgetting curve. Although Ebbinghaus has been criticized for his use of nonsense syllables, his pioneering work greatly shaped and continues to influence the field of memory research today.

Unlike Ebbinghaus, Sir Frederic Bartlett (1932) was interested in the study of meaningful information and memory. Bartlett's methodology involved having participants examine the original material (e.g., stories, folktales, pictures) and recall the original material shortly after viewing and again after varying time intervals. Bartlett's research provided early evidence for the reconstructive nature of memory. Bartlett asserted that people used their knowledge and experiences combined with elements from the original material. Although Bartlett's work was first published during the period of behaviorism and therefore received little recognition, his contributions were recognized with the rise of cognitive psychology and have influenced many memory researchers.

Other pioneers of memory research included Binet, Semon, and Ribot. Alfred Binet is well known in psychology for his work on intelligence testing. Not surprisingly, Binet researched memory as it related to the school system. Semon described the three stages of encoding, storage, and retrieval and laid the groundwork for later researchers on the encoding specificity hypothesis. Ribot investigated problems with memory. Freud's view of memory had a lasting (and controversial) impact on the field. And researchers in the stimulus-response framework, such as Pavlov, Thorndike, and Lashley, dominated the field until this framework was replaced by the information-processing approach.

Sensory Memory

Information from the environment is constantly available via the senses. For example, sights and sounds are all around. How information from the environment is registered and retained for later use is the topic of sensory memory. Sperling's (1960) classic research was on the capacity and duration of visual sensory information, referred to by Neisser (1967) as iconic memory. In Sperling's studies, he presented participants with a 3 X 4 matrix of letters (i.e., three rows with four letters in each row) for 50ms. In the first of these studies, participants were to report as many (or any) of the letters that they could remember seeing. This is called the *whole report procedure or condition*. Results indicated that participants could only remember about four or five letters—roughly a 37 percent accuracy rating.

Sperling's next step was ingenious. Because participants reported seeing more letters than what they could verbally report, Sperling developed a new condition. In this condition, participants were still exposed to the matrix of letters presented for the same length of time. However, immediately following the presentation of the matrix, participants were exposed to one of three tones (i.e., high tone, medium tone, and low tone). If participants heard the high tone, they were to report the letters on the top row. If participants heard the medium tone, they were to report the letters in the middle row. And finally, if participants heard the low tone, they were to report the letters in the bottom row. Participants had no way of knowing which of the three tones they would be given and therefore had no way of knowing which line of the matrix to focus their attention on. This is called the *partial report procedure or condition*. Results indicated that performance increased to roughly a 76 percent accuracy rating. Sperling also varied the amount of time between the matrix presentation and the tone, which signaled to participants when to report. As expected, he found that performance decreased in the partial report condition as the amount of time delay before reporting increased.

Sperling's original research has since been extended to investigate other issues in iconic memory. For example, Averbach and Coriell (1961) investigated the role of interference and found that the visual cue used (i.e., a circle marker) can interfere with the original perception. In addition, Sperling investigated the influence of the postexposure field on recall. He found that recall was significantly better when the visual field following the presentation of the matrix was dark rather than bright. Although Sperling's research on iconic memory has stimulated later research and produced a wealth of literature, it is not without critics. Specifically, Haber (1983) has stated that Sperling's studies are not ecologically valid and therefore do not reflect normal visual perception.

Sperling's research inspired others to compare the whole report condition to the partial report condition in other areas of memory. Darwin, Turvey, and Crowder (1972) did just that to investigate the capacity and duration of auditory sensory information, referred to by Neisser (1967) as *echoic memory*. In their experiments, referred to as the three-eared man, participants heard a matrix of letters and digits in the space of one second. One message (comprising three letters and digits) was played to the right ear, one to the left ear, and one to both ears (serving as the third ear). Similar to Sperling's research, where participants were visually presented with the entire matrix at one time, Darwin et al.

played all three messages at the same time. In the whole report procedure or condition, participants were instructed to report as many (or any) of the letters/digits that they could remember hearing. In the partial report condition, participants were given a visual indicator to know which message to report. Results indicated a higher accuracy rating for the partial report condition. Results from both Sperling and Darwin et al. indicate that almost complete records of visual and auditory information are available for a brief period of time after exposure. This information must be encoded (e.g., paid attention to) or it will be lost quickly.

Short-Term Memory and Working Memory

Atkinson and Shiffrin (1968) proposed a stage model of memory. In their model of memory there are three stages: (a) sensory memory, (b) short-term memory, and (c) long-term memory. According to the model, *sensory memory* is where environmental information is registered. This information fades quickly unless it is attended to for further processing (see section above on sensory memory for more information).

The second stage is *short-term memory*. Short-term memory is where new information from sensory memory is transferred and where old information from long-term memory is retrieved. This model argues that short-term memory has a limited capacity and duration. Evidence for two separate memory storages for short-term memory and long-term memory includes George Miller's (1956) classic work on the limited capacity of short-term memory (i.e., seven plus or minus two) as well as J. Brown's (1958) and L. R. Peterson and M. J. Peterson's (1959) reports where participants quickly forgot simple three-letter stimuli. Miller showed that short-term memory has a limited capacity by using a digit span task. Specifically, Miller repeated random number lists with participants immediately recalling the lists until they could no longer recall them. Results showed that regardless of the type of material used (i.e., digits, letters, etc.), people were not able to recall more than seven plus or minus two items. He did note that these items could consist of chunks of information in which multiple items are combined to form a single item. For example, instead of the letters I, R, and S taking up three item spaces, they would take up only one space as IRS. This process of combining information would later be called *chunking*.

Additionally, J. Brown and L. R. Peterson and M. J. Peterson (1959) had participants view nonsense trigrams (e.g., DCL) and then instructed participants to count backward by threes from a very large number before reporting the trigram 18 seconds later. Originally, these results were interpreted to support Atkinson and Shiffrin's short-term memory storage by demonstrating the brief nature of short-term memory (the accuracy rate was roughly 5 percent).

The third and final stage of Atkinson and Shiffrin's model is *long-term memory*. Long-term memory is considered to have an unlimited capacity and potentially permanent duration.

Atkinson and Shiffrin's model of memory is no longer the dominant model of memory, in part due to the separate stages for short-term and long-term memory. Anderson outlines the significant issues surrounding the idea of a separate short-term memory store as including (a) effects of rehearsal, (b) coding differences, and (c) the retention function. Atkinson and Shiffrin's model proposed that information was entered into long-term memory by being rehearsed in short-term memory. However, simply rehearsing information does not always improve long-term memory. For example, Craik and Lockhart's (1972) depth of processing theory asserts that rehearsal improves memory only if the rehearsal is meaningful. To illustrate this point, Craik and Lockhart showed participants a word and then had participants make three types of judgments. The shallow-level judgment asked if the word was in capital letters. The intermediate-level judgment asked if the word rhymed with another word. Finally, the deep-level judgment asked if the word fit into a sentence. Results indicated that the more deeply processed words were remembered best.

Next, a separate store was argued because of evidence suggesting that short-term memory was sensory in nature and long-term memory was semantic in nature. However, later findings indicated that sensory and semantic information can be used at short and long delays. Finally, a separate store for short-term memory was argued because of the retention function. The retention function refers to results of J. Brown (1958) and L. K. Peterson and M. J. Peterson (1959). Yet when examining these results, it must be noted that the typical pattern (with rapid initial forgetting) occurred only after participants had seen many trials. In other words, the participants were experiencing interference.

Even if a separate short-term memory storage is not required to explain research findings, information must still be rehearsed to be retained in long-term memory. Baddeley (1986, 1992) asserted that a more accurate term for the place where information is rehearsed (and worked on) is *working memory* (a term first used by Miller, Galanter, & Pribram, 1960). The term working memory implies a place where mental effort is used and conscious effort is exerted. In Baddeley's tripartite theory of working memory, there are three basic systems. The *central executive* is the main system, supporting the two slave systems: the *phonological loop* and the *visuospatial sketchpad*. Each of these components will be discussed below, including the more recently added fourth component, the *episodic buffer*.

The phonological loop processes phonological information and was proposed to account for research findings, in particular the classic digit span procedure. The phonological loop has two systems: a phonological store and an articulatory rehearsal system. The phonological store holds auditory information for about two seconds, and the rehearsal system cycles the information through a phonological store (much like subvocalization; Baddeley, 2001). Despite some critics (e.g., Neath & Nairne, 1995), the phonological loop is a simple account of data, and evidence in support of the loop includes research from patients with neuropsychological deficits and speech and language deficits (Baddeley, 2001). The visuospatial sketchpad processes visual and spatial information. There is neurological and imaging research to support the idea of the sketchpad containing both components. However, researching the two components separately has posed a challenge (Baddeley, 2001). The central executive system is involved in tasks such as decision making, reasoning, and comprehension. It is also involved in attention (e.g., focusing, dividing, and switching attention). The fourth component, the episodic buffer, serves as a third slave system. The episodic buffer processes information from long-term memory and aids in integrating that with current information in working memory.

One line of research in working memory that has been extensively investigated is how people search through items in memory when looking for a specific target—a working memory search. Although this seems like a bizarre task, it is one that is done by everyone on a daily basis. For example, imagine that a person is looking for a new release DVD at a movie rental store. He or she has seen what the DVD looks like on television and now has a memory for the DVD box. Upon arrival at the store, he or she searches the rows of DVDs, scanning each until the one that matches the record in memory is found, or not. This would be an example of a search in working memory. Generally, when one does a search, a mental representation of the target item is formed in working memory and then a search is made through a set of items until a match is found. This type of search task was developed by Sternberg (1966).

Sternberg (1966) was interested in determining how individuals search information in working memory. In Sternberg's experiment, he presented subjects with digit spans up to 6 digits long. Following the 6-digit span, Sternberg would then present participants with a single digit. This single digit would represent the target digit, and participants would have to decide if the target digit existed in the previous span. For instance, assume the digit span was 3, 4, 7, 0 and the target digit was 7. In this case, the correct answer for the subject would be yes, the target was in the digit span. If the target digit had been 2, then the correct response from the subject would have been no, the target was not in the digit span.

During this task, Sternberg recorded subjects' reaction time when it came to judging if the target was in the digit span or not. Results showed that as the digit span increased from one to six numbers, the reaction time increased as well. Sternberg proposed that participants in this task were encoding the digit span into memory, followed by encoding the target digit, and then searching through their memory to find a match. He was led to theorize that searching in working memory takes place as either a *parallel* or *serial search*. In a parallel search, the target digit is compared to all digits in the span simultaneously. This type of search predicts a flat reaction time as digit span increases. However, in a serial search, the target digit is compared to each digit in the span, one at a time in order of digit span presentation. This type of search predicts that each comparison takes a small amount of time and thereby as the digit span increases, so does reaction time. From his work, Sternberg was able to conclude that since participants' reaction time did increase with the number of digits in the span, they must have been using a serial search.

Once Sternberg (1966) had concluded that participants were engaging in a serial search, he questioned if the serial search was exhaustive or self-terminating. In an exhaustive search, participants are searching the digit span in memory for the target digit through the whole digit span. Sternberg indicated that this would occur when the target digit was not in the digit span. For example, if the digit span consisted of 4, 2, 6, 7, 0 and the target digit was 9, then the participant would have to search the whole span in order to know that the target was not in the span. In a self-terminating search, once the target digit is found in memory, the search stops. Therefore, if the digit span was 4, 2, 6, 7, 0 and the target digit was 6, the search would terminate after reaching the third digit in the span. This would result in a far shorter search as compared to the exhaustive search. Sternberg believed that a self-terminating search would occur only when the target digit was present in the digit span. However, when comparing reaction time for a target digit that was in the digit span and for one that was not, the reaction time was the same. This indicated not only that participants were using a serial search in working memory, but also that the search was always exhaustive.

Long-Term Memory

The final stage to Atkinson and Shiffrin's (1968) model is the concept of long-term memory. The purpose of long-term memory is to organize and store information in memory. Long-term memory involves the processes of acquisition, retention, and retrieval of information. Most researchers believe that long-term memory is unlimited in its capacity to store and organize information. Additionally, most researchers believe that once that information is stored and organized in long-term memory, it will be kept permanently. Squire distinguished between two main divisions in long-term memory. Specifically, long-term memory can be divided into declarative (i.e., explicit) and nondeclarative (i.e., implicit) memory.

The first subdivision of long-term memory is declarative memory. *Declarative memory* is memory that requires a person to consciously recall information. Declarative memory can be divided into either semantic or episodic memories. *Semantic memory* is one's memory for general knowledge and facts. For example, knowing that a car has four tires or that the first president of the United States of America was George Washington are examples of using semantic memory. On the other hand, *episodic memory* is a memory for episodes in one's life—for example, knowing the date of one's own birthday or knowing what one ate for lunch yesterday. Often, the term episodic memory is used interchangeably with the term *autobiographical memory*.

Autobiographical memory, the memory for events in a person's life, has been used as a tool to explore many areas of interest in psychology (Conway & Pleydell-Pearce, 2000). For example, autobiographical memory has been used to study distressing personal events (Chapman & Underwood, 2000) and the lives of famous people (Neisser, 1981). Researchers have also investigated changes in autobiographical memory across one's lifespan, from infantile amnesia to changes in the function of memory as people age. Studies have identified some general characteristics of autobiographical memory including the following: They are experienced as having occurred at a unique time, being experienced by the self, containing visual imagery, being highly affect laden, and being remembered as highly accurate (Leahey & Harris, 2001). Other important issues of concern, such as the stability and accuracy of the autobiographical memory, are also examined in these studies.

To address these issues of concern and to develop this area of research in human memory, researchers have developed many techniques for studying autobiographical memory. For example, one commonly used technique involves having participants keep a diary to record their daily events and later testing participants over the contents of the diary. Another technique is to have participants record an event that is occurring when they are contacted by beeper at random time intervals (Brewer, 1988).

One particularly interesting type of autobiographical memory is a flashbulb memory (Brown & Kulick, 1977). A *flashbulb memory* occurs when a vivid, rare, or significant personal event takes place, usually within the context of hearing about a major (typically tragic) public event. Unlike much memory research, which can be performed within a laboratory setting, research on flashbulb memory cannot. Typically, when a significant public event occurs, researchers must act quickly to take advantage of this opportunity to examine flashbulb memories. For example, Talarico and Rubin (2003) studied people's flashbulb memory following the events of the 9/11 terrorist attacks. Specifically, the researchers tested undergraduates' memory the day after the events and then again 1, 6, or 32 weeks later. At each session, Talarico and Rubin asked the participants to record their memories of when they first heard of the attacks as well as a recent "everyday" memory. Results from this study showed that as time passed, the consistency in their flashbulb and "everyday" memories declined. However, the researchers also found that as time passed, the participants believed that their flashbulb memories were highly accurate. Thus, the researchers concluded that although flashbulb memories are not remembered any better than "everyday" memories, a person's confidence in them is generally much higher.

However, not all research on flashbulb memories has arrived at the same result. A study conducted in England in the early part of the 1990s produced a different result. Researchers in England studied the impact of Margaret Thatcher's sudden resignation in 1990. Using a similar methodology as in the aforementioned studies, Conway (1995) found that participants who were more invested in the event were more consistent in their recall of their flashbulb memories over time. Therefore, the level of importance attached to the event for the individual in the study appeared to be the determining factor when evaluating consistency. In other words, people are more likely to form flashbulb memories if they view the event as especially important to them and it has an emotional effect.

Recently, researchers used autobiographical memory to investigate memories of media experiences. Hoekstra, Harris, and Helmick (1999) asked young adults to recall from their childhood a movie that had seriously frightened them. Almost all participants (98 percent) had such a memory and could describe the experience and its effects vividly, reporting sleep disturbances, specific and nonspecific fears, and preoccupation with stimuli from the film. Later research used autobiographical memory to look at memories for frightening movies seen on dates (Harris et al., 2000). Sex differences existed in the behaviors participants engaged in while viewing the movie (e.g., women reported being more jumpy than men). Another study (Harris, Hoekstra, Scott, Sanborn, & Dodds, 2004) had participants report their memories for watching romantic movies on dates. Sex differences were found in who originally chose to watch the movie, behaviors while watching the movie (e.g., men reported laughing less than women), and cognitions engaged in while watching the movie (e.g., men reported thinking about their date more than women did). Autobiographical memory has also been used as a tool to study other areas in media psychology, including televised sporting events (Bonds-Raacke & Harris, 2006), portrayals of gay and lesbian characters (Bonds-Raacke, Cady, Schlegel, Harris, & Firebaugh, in press), and memories for music. See Harris, Bonds-Raacke, and Cady (2005) for a complete review of using autobiographical memory to study media.

APPLICATIONS

Over the years, memory has been researched extensively and along many different lines. Many of these lines have provided psychologists with answers to the structure of memory, how retrieval is affected, and ways to enhance one's memory. Thus, this section will discuss models of memory, forgetting, factors affecting recall, and ways to improve memory.

Models of Memory

With the advent of the cognitive revolution, a wealth of research on memory has developed. Much of this research has evaluated how information is organized, stored, and recalled from memory. Specifically, researchers have developed models of memory to account for data generated in the lab. These models provide a systematic framework for formulating predictions, thereby supplying researchers with a way to connect research in the laboratory with events in the real world. The most popular models of memory include the SAM model (Raaijmakers & Shiffrin, 1981), the Neural Networks model, and Anderson's (1983) ACT model.

The SAM model is a broad-based mathematical model of memory that seeks to account for results obtained using list learning paradigms. The idea here is that words have a certain memory strength or familiarity value. This strength is based on three cues: (a) the context in which the item is encoded, (b) other items to be remembered, and (c) the item itself. These three cues will determine the strength of the word in one's memory. The latter two cues are dependent on the number of rehearsals an item receives during the encoding process. The greater the rehearsal number, the stronger the cues and the greater the memory strength will be. As with any memory task, the context in which the item is remembered is also important to recall. Therefore, the memory strength is mediated by the context in which the encoding took place. Additionally, in the SAM model, it is these three types of cues that will either facilitate or inhibit later retrieval of information. However, the SAM model has been criticized in the past for having too many assumptions, not all of which can be empirically tested (Roediger, 1993).

A Neural Network model of memory is different from the mathematical model of SAM. A Neural Network model is a computer simulation of neurons in the brain. Each neuron in the network receives input from the environment or nearby neurons and then produces an output as represented by an activation function. In a person, input information occurs through physical energy (i.e., light), whereas output information is expressed through mechanical energy (i.e., movement). In a neural network, these types of energy are represented by activation of either an excitatory or an inhibitory nature. Activation levels are generally expressed as a number from 1.0 to –1.0.

Each neuron is connected with other neurons by links that are excitatory or inhibitory. These networks of words and concepts vary in complexity from the simple to the extreme. The strength of the links between neurons in this model is determined by weights. Weights are a direct function of the neurons' learning history. The more an item (neuron) is encoded and used, the stronger the weight. Therefore, the output of a neuron in a Neural Network model is determined by the activation of an input neuron and the weights attached to the underlying links for that given neuron. If the weights are strong, then the links will be activated and additional neurons excited or inhibited. If the weights are weak, then no information will be outputted. This theory is very popular among applied fields due to its mathematical nature and its ability to learn new information and make predictions (Massaro, 1988).

The final model of memory is Anderson's (1983) ACT model. Unlike many models that focus solely on declarative memory, the ACT model evaluates declarative, procedural, and working memory. This model also explicitly evaluates the impact of the environment, whereas other models (i.e., SAM model) do not. The emphasis with ACT has been to model the acquisition of problem-solving skills.

Each of the three aforementioned types of memory is represented by ACT. Declarative memory provides the model with information about facts and general knowledge; this is where permanent memory resides. In ACT, the procedural memory holds production rules, which are either weakened or strengthened by their fit to environmental demands. Finally, as with previous research on working memory, this type of memory is the workhorse of ACT. Information from the environment requires information to be recalled from declarative and procedural memory. The structure in the ACT model that allows for the integration of this information (i.e., knowledge and production rules) and the changing of this integration as a result of environmental influence is working memory. Due to this flexibility, ACT has been very popular and has accounted for results in skill acquisition (Anderson, 1983), processing in working memory, and problem solving.

Forgetting

Typically, people think about their memory not when trying to learn information but rather when memory has failed them. When one's memory fails, the most common reason cited by people is that they have forgotten some piece of information. Forgetting is defined as a loss of memory. The interesting question is, why does this event occur? Once information is learned, the rate of forgetting is rapid. Ebbinghaus showed this in his work with the forgetting curve. This idea that information simply fades with time is known as the *decay theory*.

However, time is not the only factor in forgetting. Retention of information is also affected by interfering material. McGeoch (1942) distinguished between two different types of interference. The first type of interference is called retroactive interference. *Retroactive interference* occurs when new information disrupts previously acquired information. For example, after moving to a new home, most people will get a new phone number. After a few years, many people find it hard to recall their previously held phone number. This is an example of retroactive interference in that the new phone number disrupts the previously held phone number. The second type of interference is proactive interference. *Proactive interference* occurs when previously learned information disrupts the acquisition of new information. For example, when learning information for a new psychology class, some people struggle with information from other psychology classes interfering with learning the new information. This would be an example of previously learned information disrupting the retention of newly learned information.

Additional research on interference has looked at the fan effect (Anderson, 1974). The *fan effect* describes a function in which reaction time on a recall test increases in proportion to the number of facts associated with a concept. This can be explained by some of the models of memory in that spreading activation can cause interference. Specifically, when a particular piece of information about a concept is to be recalled, spreading activation with that concept will determine the speed of recall. If other pieces of information have higher strength, that piece of information will cause interference and slower recall. Thus, the more facts associated with a concept, the slower the retrieval of any one of the facts.

Therefore, there appear to be two separate mechanisms that produce forgetting: decay and interference. However, many psychologists suggest that decay of information really does not occur. Rather, psychologists have suggested that all forgetting is a function of interference. Specifically, although memories appear to decay over time, it is more likely that memories are interfered with by new memories that participants have encoded.

Factors Affecting Recall

In addition to research on forgetting, there is a plethora of information evaluating other factors that affect recall from memory. One of the best-known factors that affect recall is serial position. When conducting a recall task, there are two types of recall a researcher can request. The first is known as *free recall*. Here, participants can recall items in a list in any order they choose. The second is known as *serial recall*. In this task, the participants must recall items in a list in the correct order of presentation. In this task, the participants must rehearse not only the item itself but also the position of the item in the list. This added level of complexity to the task makes it more difficult than the free recall task. As the researcher shows more and more items, the participants tend to perform more and more poorly on the task. Researchers consistently see the same curve during this task. Specifically, participants have better recall for those items at the beginning and the end of the task list and extremely poor recall for items in the middle; this pattern is known as the *serial position effect* or *serial position curve*.

Two separate events, known as *serial position effects*, are happening to create the serial position curve. First, participants have accurate recall for items presented at the beginning of a list. This accuracy for early information is called the *primacy effect*. If participants have a strong primacy effect, then they are engaging in good rehearsal over the earlier items. This good rehearsal is increasing the likelihood that information is moved to long-term memory—for example, remembering the first few items on a grocery list but not the rest. However, if participants have a weak primacy effect, then they are engaging in insufficient rehearsal of early list items and decreasing the likelihood of items moving to long-term memory.

Second, participants have extremely accurate recall for items near the end of the list. This accuracy on the final items on the list is known as the *recency effect*. Research has shown that most people have a strong recency effect due to the capacity of short-term memory (i.e., 7 +/–2 items). Not too surprisingly, in order to eliminate the recency effect researchers provide a distracter (similar to L. R. Peterson & M. S. Peterson, 1959) task following the last item in the list to be remembered. Glanzer and Cunitz (1966) applied a distracter task, a counting task, following a serial position task. Results indicated that participants had a weak recency effect as recall was delayed, indicating the impact of short-term memory on late item recall. However, the primacy effect was unaffected by this task, supporting the moving of earlier items into long-term memory via rehearsal.

The *physical context* of learning information, such as where a person was during encoding, can impact a person's recall ability at a later point by providing cues. However, not all research has demonstrated the validity of this effect, and researchers have argued that the likelihood of this effect's taking place depends on the degree to which participants integrate the context with their memories. There are several other types of context effects, including emotional context, mood congruence, and state-dependent memory.

Like physical context, *emotional context* has been shown to either improve or hinder recall. However, emotional context effects appear to occur only under special circumstances.

A stronger effect is witnessed with *mood congruence*. Unlike the physical context and emotional context effects, mood congruence does not occur because of the context of learning. Rather, participants are presented with positive, neutral, and negative words to learn. At the time of recall, researchers induce either a positive or a negative state. Results show participants recall more of the words that match their moods.

Related to mood congruence is the concept of *state-dependent memory*. State-dependent memory extends the concept of context to the internal state of the individual. Specifically, research has shown that people are better at recall when their internal state at test is the same as their internal state during encoding. Interestingly, much research on state-dependent memory has looked at the effects of alcohol and drugs on recall of information.

Improving Memory

Although it has been discussed indirectly, much research has focused on improving people's ability to retain and retrieve information from their memories, specifically by focusing on ways to enhance the encoding of information. Examples of this include (a) elaborative rehearsal, (b) depth of processing, and (c) the use of mnemonic devices.

Earlier in the chapter, we mentioned the most basic type of rehearsal: maintenance rehearsal. *Maintenance rehearsal* is a low-level repetitive recycling of information. This is the type of rehearsal one would use when remembering a phone number. However, in terms of aiding encoding and memory, once that recycling stops, there is no permanent memory. Yet, Craik and Tulving (1975) identified another type of rehearsal. This type, *elaborative rehearsal*, is more complex in that it uses the meaning of the information for storing and remembering. When one uses elaborative rehearsal, information is encoded more deeply and thus has a greater probability of being retained.

These two types of rehearsal were used to explain Craik and Lockhart's (1972) depth-of-processing model. As mentioned earlier, when information is presented to be encoded, information is encoded at different levels. Some information will receive little attention and little rehearsal and therefore will not be encoded very deep. Other information will be subjected to more intentional and elaborative rehearsal and deeper, more meaningful processing. It is these types of processing that will determine the likelihood that information can be recalled at a later date.

Interestingly, another way to improve memory through improving encoding also has to do with the level of processing. Many people use mnemonics to encode information at a deeper level so that recall will improve. The term *mnemonic* means "helping memory." A mnemonic device is one where strategic rehearsal is employed. The strength of mnemonics is based on the fact that information is repeatedly rehearsed, the information is built into existing frameworks, and the mnemonic device produces an enhanced ability to recall information.

One of the earliest developed and known mnemonic devices was the *method of loci*. This technique uses a combination of mental imagery and physical locations when encoding information. When using this method, a person will come up with a list of 10 to 15 easily remembered locations. For example, one could develop a list based on locations passed on the drive home from school or work. A person would take the first item to be remembered and mentally place that item in the location. Then, the second item to be remembered would be placed it in the second location, and so on. When the time comes for the information to be recalled, the person can just mentally take a drive and go "looking" for the items of information placed in each location. Research has shown that the distinctiveness of the mental image will aid in the recall of information.

The second most popular mnemonic device is that of the *peg-word method* (Miller, Galanter, & Pribram, 1960). In the peg-word method, a prememorized set of words serves as a set of mental "pegs" onto which the information to be remembered can be "hung." The peg-word method relies on rhymes with the numbers one through ten. For example, in the peg-word method, one is a bun, two is a shoe, three is a tree, and so on. The information that is to be remembered is then "hung" on the "pegs" with the person forming a mental image of the rhyming word and the information to be remembered. For example, if the information to be remembered is a grocery list and the first item is ketchup, a person would make a mental association between "one is a bun" and ketchup. Therefore, a person could picture a bun being smeared with ketchup. Again, the distinctiveness of the mental image will aid in the recall of information.

SUMMARY

The topic of memory has been discussed and studied for centuries. Although the topic itself is relatively old, much of the empirical research findings have come about only in the last 125 years or so. Space limitations preclude us from discussing all memory topics. However, other chapters in this Handbook present such topics as false memories and eyewitness testimony. Finally, even though this chapter and psychologists have just scratched the surface of human memory, the information gathered thus far has led to a wealth of knowledge explaining how people encode, store, and retrieve information.

REFERENCES AND FURTHER READINGS

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451–474.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal, 40,* 309–328.

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255,* 556–559.

Baddeley, A. D. (2001). Is working memory still working? *American Psychologists, 56,* 849–864.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Bonds-Raacke, J. M., Cady, E. T., Schlegel, R., Harris, R. J., & Firebaugh, L. (in press). Remembering gay/lesbian television and film characters: Can Ellen and Will improve heterosexuals' attitudes towards homosexuals? *Journal of Homosexuality*.

Bonds-Raacke, J. M., & Harris, R. J. (2006). Autobiographical memories of televised sporting events watched in different social settings. *Psychological Reports, 99,* 197–208.

Brewer, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 21–90). Cambridge, UK: Cambridge University Press.

Brown, J. (1958). Some tests of decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 19,* 12–21.

Brown, R., & Kulick, J. (1977). Flashbulb memories. *Cognition, 5,* 73–99.

Chapman, P., & Underwood, G. (2000). Forgetting near-accidents: The roles of severity, culpability and experience in the poor recall of dangerous driving situations. *Applied Cognitive Psychology, 14,* 31–44.

Conway, M. A. (1995). *Flashbulb memories*. Hillsdale, NJ: Erlbaum.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self system. *Psychological Review, 107,* 261–288.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104,* 268–294.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). The auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3,* 255–267.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

Haber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences, 6,* 1–11.

Harris, R. J., Bonds-Raacke, J. M., & Cady, E. T. (2005). What we remember from television and movies: Using autobiographical memory to study media. In W. R. Walker & D. J. Herrman (Eds.), *Cognitive technology: Essays on the transformation of thought and society* (pp. 130–148). Jefferson, NC: McFarland & Co.

Harris, R. J., Hoekstra, S. J., Scott, C. L., Sanborn, F. W., & Dodds, L. A., (2004) Autobiographical memories for seeing romantic movies on a date: Romance is not just for women. *Media Psychology, 6*(3), 257–284.

Harris, R. J., Hoekstra, S. J., Scott, C. L., Sanborn, F. W., Karafa, J. A., & Brandenburg, J. D. (2000). Young men's and women's different autobiographical memories of the experience of seeing frightening movies on a date. *Media Psychology, 2,* 245–268.

Hoekstra, S. J., Harris, R. J., & Helmich, A. L. (1999). Autobiographical memories about the experience of seeing frightening movies in childhood. *Media Psychology, 1,* 117–140.

Leahey, T. H., & Harris, R. J. (2001). *Learning and Cognition* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

McGeoch, J. A. (1942). *The psychology of human learning.* New York: Longmans.

Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart, & Winston.

Neath, L., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace-decay theory. *Psychonomic Bulletin and Review, 7,* 403–423.

Neisser, U. (1967). *Cognitive Psychology.* New York: Appleton.

Neisser, U. (1981). John Dean's memory: A case study. *Cognition, 9,* 1–22.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). SAM: Search for associative memory. *Psychological Review, 88,* 93–134.

Roediger, H. L., III. (1993). Learning and memory: Progress and challenge. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 509–526). Cambridge, MA: Bradford/MIT.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*, 45.

47

Memory and Eyewitness Testimony

Kerri Pickel
Ball State University

According to the Innocence Project (*www.innocenceproject.org*), created by Barry Scheck and Peter Neufeld, approximately 75 percent of the wrongly convicted people who have been eventually freed from prison through forensic DNA testing were found guilty at trial because of mistaken eyewitness identifications. The results of many studies have demonstrated that witnesses can make errors when they try to describe perpetrators or identify them in lineups. However, jurors tend to trust witnesses' accounts and give them substantial weight when making decisions about defendants' guilt. Studying the variables that affect the accuracy of eyewitness memory is important because the knowledge gained could lead to better techniques for questioning witnesses and more effective evaluations of witnesses' statements.

METHODOLOGY

In the 1970s, Elizabeth Loftus began publishing groundbreaking studies showing that questioning witnesses in certain ways can change their reports. Her work inspired numerous other researchers to examine various aspects of eyewitness performance. Most eyewitness memory studies are laboratory experiments, which have the advantage of allowing researchers to control extraneous variables and confounds and thus improve the chances of discovering causal relationships. The participants are often college students, although sometimes community members are recruited. The participants act as witnesses by viewing a video, a slide sequence, or a live, staged event enacted by confederates. The researchers may disguise the purpose of the experiment so that, while viewing the critical event, the witnesses are unaware that they will later have to try to remember it. In this way, researchers hope the participant-witnesses' experience will be similar to that of actual witnesses, who are often simply going about their daily activities when they are surprised to notice a crime unfolding in their vicinity.

After watching the critical event, the witnesses may complete a written questionnaire that asks them to remember what happened and to describe the characteristics of the perpetrator. Alternatively, they may be interviewed by the experimenter, with the interview recorded for later analysis. In addition to or instead of asking for descriptive information about the event or the perpetrator, the witness could be required to view a lineup in which they try to identify the guilty person. Most lineups are photo lineups, in which witnesses examine a set of mugshot-type photos, but it is also possible to use a live version so that the lineup members are physically present. As in the real world, the true perpetrator may or may not appear in the lineup along with the innocent distracters (called fillers or foils). If he or she does, it is a *target-present* lineup. When a witness correctly identifies the target, it is called a *hit*. *Target-absent* lineups are created in the lab by replacing the perpetrator with a foil who resembles him or her. A *false alarm* occurs when a witness incorrectly identifies a foil as the perpetrator. As the Innocence Project has shown, it is this error that can send individuals to prison for crimes they did not commit.

In a laboratory experiment, the researcher can manipulate one or more independent variables to observe the effect on the witnesses' memory. The independent variable could relate to some aspect of the critical event (e.g., whether the perpetrator is carrying a weapon), the wording of the questions asked after the event, or the way the lineup is constructed. In this way, researchers can try to determine whether these variables causally influence memory. Although experiments can be quite effective at accomplishing this goal, they could be criticized for having low external validity. In other words, concerns could be raised about whether the results would generalize to actual crime situations due to certain characteristics of the experiments. For example, college student participants may differ in important ways from typical witnesses to real crimes, experimental witnesses may feel less anxiety during the critical event, and they know that their reports will have no real consequences.

An alternative research design that obviously has higher external validity is the field study, which typically uses an archival approach to analyze the reports or lineup identifications made by real witnesses. Although the external validity is good, there are two related disadvantages of this type of design. First, there are usually multiple uncontrolled variables that can make it difficult to determine cause and effect. For example, across cases there may be variations in the duration of the crime event, the physical distance between the witnesses and the perpetrators, the elapsed time between the event and the witness's report, the interviewing techniques used by police, and lineup instructions. Second, when studying actual crimes it is difficult (and sometimes impossible) to figure out exactly what happened, who committed the crime, and how the perpetrator looked at the time, and this information is clearly needed to assess the accuracy of witnesses' reports. Even when someone is arrested and convicted, researchers cannot be sure that this person was the perpetrator, as the Innocence Project's data demonstrate. Moreover, when it can be established that a certain person is guilty, it is still hard to know exactly what he or she did during the crime or whether the witness provided an accurate description of his or her appearance at the time.

Because lab and field studies have different strengths, they can be seen as complementary, and researchers can use both to learn about eyewitness memory. Together the two types have revealed a number of variables that affect witness accuracy. Researchers have found it convenient to classify these variables into two categories: *system* and *estimator variables*. The former are so named because they can be controlled by the criminal justice system. For instance, police investigators can choose the instructions that they give to witnesses preparing to view a lineup. Estimator variables can be estimated but not controlled by the system. An example of this type of variable would be whether the witness happens to be of the same race as the perpetrator. It is not within the scope of this article to describe in detail all of the many variables that researchers have linked to eyewitness performance, but the next sections will review what is known about some in each category.

ESTIMATOR VARIABLES

Effects of Stress

It seems logical to assume that witnessing a crime would be stressful, particularly if the perpetrator has a weapon or commits violence. It is difficult to know for sure how witnesses feel during a real crime, however, because the best that researchers can do is to ask these witnesses to self-report their emotional state at some time after the event. Retrospective reports can be inaccurate because of witnesses' imperfect memory of their prior state (especially after longer delays) and biases such as understating of the level of anxiety felt during the crime due to a motivation to minimize the dangerousness of the experience. Despite the measurement problems, researchers have asked real-life witnesses to rate their anxiety during the crime. The findings show differences among witnesses, some of which can be attributed to various aspects of the crime situation, such as the witnesses' closeness to the perpetrator and the duration of the event. Not all witnesses report feeling terrified. Nevertheless, some clearly experience more stress than they would during a normal daily event, and therefore it is important for researchers to explore the effects of stress on memory accuracy.

A starting point is to define what is meant by stress. Sometimes the terms *stress*, *anxiety*, *arousal*, and *negative emotional stress* are used almost interchangeably within the literature, and these terms confuse physiological and emotional reactions. However, researchers have begun to distinguish between two types of responses to stimuli that are ostensibly stressful. One is a *defensive* physiological response that occurs when the *activation mode* of attentional control dominates. It involves increases in heart rate and blood pressure as well as vigilance and readiness for action (e.g., escape). This type of response can be triggered by stimuli that cause cognitive anxiety or worry as well as the perception of physiological arousal. The second type is an *orienting* response that occurs when the *arousal mode* of attentional control is more active. Here, heart rate and blood pressure are lowered, and attention is directed to stimuli that seem informative or novel to the individual. These stimuli could have negative emotional aspects to them, but not necessarily. For the present purposes, a stressful stimulus is one that produces a defensive response, and a negative emotional stimulus will be defined as one designed to cause witnesses to feel some sort of negative emotion (e.g., sadness, revulsion), probably without producing a defensive response, as discussed below.

At one time researchers generally argued that stress causes witnesses to allocate their attention to the "central" details of a scene (details connected with the theme or gist of the event) at the expense of "peripheral" details (irrelevant to the gist). For example, in one representative lab experiment, Christianson and Loftus (1991) asked witnesses to watch a slide sequence in which the critical slide showed either a neutral event (a woman riding a bicycle) or a negative emotional event (the aftermath of a car-bicycle accident in which the woman was injured and shown lying on the ground, bleeding from a head wound). The witnesses subsequently remembered central information from the critical slide (the color of the victim's coat) more accurately if the slide was emotional rather than neutral, but the reverse was true for the peripheral information (the color of an uninvolved car in the background). Results like these were usually explained by referring to the Yerkes-Dodson law and J. A. Easterbrook's (1959) cue-utilization hypothesis. According to the Yerkes-Dodson law, stress predicts performance as an upside-down U function, so that memory accuracy declines as stress increases above the optimal level in the center of the curve. Easterbrook's hypothesis was used to explain in more detail the cognitive processes that occur: As stress increases, the range of stimuli that an individual can attend to becomes narrower, and consequently the individual must allocate his or her attention to the most informative or useful stimuli in the environment (i.e., the central details) instead of other stimuli (i.e., the peripheral details).

More recently, some researchers have reinterpreted the studies that seemed to demonstrate that stress improves memory for central details while harming memory for peripheral information. In their review, Deffenbacher, Bornstein, Penrod, and McGorty (1994) argued that these studies elicited an orienting response rather than a defensive response. In other words, they thought that, although the emotional stimuli attracted the witnesses' attention (perhaps because of their salience, unusualness, or importance to the story), they did not actually cause stress. If it is true that witnesses allocated considerable attentional resources to the emotional stimuli, it would make sense that they later remembered less about other details than they otherwise would have.

Obviously, to determine the effects of stress on memory, one must look at studies in which stress was truly manipulated or varied naturally across conditions. Deffenbacher et al. (1994) contended that such studies do exist, but they have sometimes been confused with ones that merely include negative emotional stimuli. A set of studies by Douglas Peters (1988, 1991, 1997) serves as an example. In them, the stressful stimulus was a medical inoculation that participants received. In a meta-analysis that included Peters's studies and others, Deffenbacher et al. reported that high levels of stress impaired both lineup identification and descriptions of target persons (if measured by answers to specific questions asked during an interview but not free recall). With respect to lineups, the authors speculated that high stress might degrade the witness's memory representation of the perpetrator so that it becomes difficult to match that representation to him or her in a target-present array but easy to see similarities between the memory representation and one of the foils in a target-absent array. Deffenbacher et al. also suggested that the impaired memory representation of the target might be less apparent when free recall is used instead of interrogation because in the former case the witness can decide what to include, what to exclude, and in what order to report the information.

Exactly how are stress and memory accuracy related? To answer this question, Deffenbacher et al. (1994) appealed to a "catastrophe model." In contrast to the Yerkes-Dodson curve, which predicts that as stress increases above the optimal level memory performance will gradually drop, the catastrophe model proposes that, at some critical point, the drop will be sudden and discontinuous if stress reactions have become too extreme. As a result, memory for the stressful event will be weak.

Biologically, stress causes the adrenal glands to release glucocorticoids. If their levels are high enough, these hormones can disrupt the normal functioning of the hippocampus, which is a brain structure instrumental in organizing and consolidating episodic details into coherent long-term memories. Although some aspects of the overall gist of a stressful event may be reasonably well remembered thanks to the involvement of another brain structure called the amygdala, the memory representation may lack spatial, temporal, and other contextual details if the hippocampus is not working normally. Given the discussion thus far, it may seem puzzling that individuals who have experienced an extremely stressful event (one we might call traumatic) often have memories of that event that are vivid and intrusive, so that unwanted flashbacks may occur. However, traumatic memories are not necessarily accurate, and sometimes details of one event are blended with those of another similar experience, or the account includes details that did not happen at all but instead were simply imagined or thought about. In addition, traumatic memories are often fragmented and lacking in contextual detail, as when a soldier recalls mainly a mental image of a dead comrade's face when remembering a combat episode.

In sum, empirical evidence suggests that negative emotional stimuli most likely attract attention by causing an orienting response, but witnesses probably do not experience stress when viewing them. These stimuli will receive favored processing relative to neutral stimuli, but the advantage is restricted to central details related to the emotional stimuli as opposed to peripheral or background information. Stressful stimuli, on the other hand, seem to produce a defensive response that results in a general impairment of memory except perhaps the overall gist.

The Weapon Focus Effect

Witnesses who observe an armed perpetrator tend to direct their attention toward the weapon, and subsequently they are unable to remember information about the perpetrator's physical appearance as accurately as they would have if no weapon had been visible. This phenomenon is known as the *weapon focus effect.* Many experiments have demonstrated that a weapon's presence reduces the accuracy of witnesses' descriptions of the perpetrator and their ability to identify that person in a lineup, although the effect size with the latter variable is weaker, probably because it is a less sensitive measure of memory.

A good illustration of the effect is found in a study conducted by E. F. Loftus, G. R. Loftus, and Messo (1987). Their witnesses viewed a slide sequence depicting customers moving through the order line at a fast-food restaurant. The perpetrator, who was shown on four critical slides, held either a gun or a personal check as he took money from the cashier. Witnesses who saw the gun were less likely than those who saw the check to identify the perpetrator in a photo lineup (in two experiments), and their descriptions of him were less accurate (in one experiment). E. F. Loftus et al. used a corneal reflection device to track witnesses' eye movements as they watched the slides. They found that the witnesses fixated on the gun more frequently and for longer durations than they fixated on the check. This result, along with the finding that the weapon did not affect witnesses' statements about details presented before the critical slides, implies that weapons both attract attention and decrease witnesses' ability to encode other details that appear simultaneously with the weapon.

In addition to E. F. Loftus et al.'s (1987) study, many other lab experiments have obtained the weapon focus effect. Various types of stimuli have been used (slides, video, and live, staged events), as well as different weapons (e.g., handgun, knife, meat cleaver, liquor bottle, and syringe). Furthermore, community samples as well as college students have participated.

Among field investigations of the effect, the findings are mixed, with many (but not all) reporting null results. Remember, field studies are complicated by the presence of uncontrolled variables that could influence witnesses' reports and obscure the causal relationship between the weapon and memory for the perpetrator, if this connection exists. For example, an armed criminal might feel emboldened to approach nearer to the witness or stay at the scene longer than an unarmed one, which in turn could cause the witness to encode information about the criminal more completely and more deeply, thus mitigating the weapon focus effect. Similarly, the police might regard crimes involving a weapon to be more serious, so they might question witnesses to these crimes more thoroughly and therefore elicit better descriptions. More research is needed to clarify the field results.

Among the possible explanations for the weapon focus effect, the most intuitively appealing may be that a weapon causes witnesses to feel frightened or anxious due to the possibility that they or someone else could be injured or killed. As their anxiety rises, their attentional capacity shrinks so that they become unable to encode as many details simultaneously as they normally would. Therefore, they focus on details that are considered important, such as the weapon, which is the source of the anxiety. Less important details, such as the color of the perpetrator's shirt, receive minimal processing. This explanation, which is based on Easterbrook's (1959) cue-utilization hypothesis, is opposed by several results. First, the weapon focus effect has been obtained even when witnesses' rated anxiety levels are low. This result suggests that an orienting response rather than a defensive response may be at work, and that weapons do not necessarily cause genuine stress. Additionally, witnesses' memory for the perpetrator is not harmed by the degree to which the perpetrator threatens a victim or the degree of threat associated with the object held by the perpetrator.

A better explanation is based on unusualness: Weapons are unexpected within many contexts. For example, even though shoppers know that convenience stores sometimes get robbed, the probability of witnessing a robbery (and seeing a gun) during any given visit to one of these stores is low. Moreover, previous attention research has shown that people look more often and for longer durations at unusual rather than typical objects within visual scenes. In support of this hypothesis, researchers have found that an unusual object (e.g., a stalk of celery or a raw chicken) can reduce memory accuracy just as a weapon can. Also, in accordance with the hypothesis, the weapon focus effect disappears if the witness sees the weapon in a context in which it would be expected such as a shooting range.

Recently researchers have tried to determine whether witnesses' fixations on a weapon are automatic (and thus involuntary and unavoidable) or controllable. This issue is interesting for both theoretical reasons, because it has implications for researchers' ability to understand witnesses' attentional processes, and practical reasons, because it speaks to the possibility of training potential witnesses to provide more accurate reports. Pickel, Ross, and Truelove's (2006) participants listened to a lecture that either informed them about the weapon focus effect, instructing them to try to avoid focusing on a weapon, or covered an unrelated topic. Subsequently the participants witnessed a staged event involving a perpetrator who carried either a gun or a textbook. The witnesses who heard the control lecture exhibited the usual weapon focus effect, describing the perpetrator more accurately when he held the book instead of the gun. In contrast, the educated witnesses remembered the perpetrator equally well in the two object conditions. From these results, Pickel et al. concluded that weapons probably do not capture attention automatically, and witnesses can overcome the effect if informed about it.

The Own-Race Bias

Witnesses can more accurately remember faces of members of their own racial group than those of another group. Meissner and Brigham (2001) conducted a meta-analysis that reviewed 91 different studies in an effort to find overall statistical conclusions about this effect. They discovered that the own-race bias is apparent when both hits and false alarms are used as dependent measures. Across the studies they reviewed, witnesses were 1.4 times more likely to make a correct lineup identification if the target was of the same rather than a different race, and they were 1.56 times more likely to make a mistaken identification. They also discovered that, although witnesses of any race can display the own-race bias, the effect was larger for white than for black witnesses when considering false alarms (but not hits). Furthermore, Meissner and Brigham reported that false alarms for other-race faces increased when witnesses viewed the target for a shorter time. Additionally, follow-up research has demonstrated that witnesses are highly confident in their mistaken identifications of other-race targets and that children as young as kindergarten age show the same bias that adult witnesses do.

Several social and cognitive processes have been invoked to interpret the own-race bias. One proposal is that witnesses with prejudiced racial attitudes are not very motivated to differentiate between other-race individuals. However, most recent studies have not obtained a correlation between attitudes and other-race memory performance. Researchers have also examined the possibility that the faces of members of some racial groups might show less physical variability than others, which would make those faces hard to distinguish. As with the first proposal, little support was found for this hypothesis. The most promising explanation of the other-race bias draws upon results showing that witnesses who have had more contact with members of other races are less susceptible to the effect. For example, in one study, children and adolescents living in integrated neighborhoods could better recognize other-race faces than those living in segregated areas could. Moreover, several data sets have indicated a positive correlation between witnesses' self-reported level of contact with other-race individuals and the accuracy of their lineup identifications of other-race targets (Meissner & Brigham, 2001).

Assuming that interracial contact matters, the next step for researchers is to determine why. Although decreases in prejudice and stereotypic beliefs could play a role, most accounts focus on perceptual learning, or the idea that increased contact helps one to learn more about the physical features associated with other-race faces and to memorize them better. Specifically, researchers have suggested that, with experience, individuals learn to identify and attend to the features that are most relevant to the task of discriminating one face from another. Over time, other-race faces can be encoded in memory more efficiently and with a richer, more complete representation.

Even with the possibility that individuals may be able to reduce their susceptibility to the own-race bias through interracial contact, the effect remains problematic for the legal system. As Meissner and Brigham (2001) noted, it is especially troubling that mistaken identifications are more likely if the witness and target are of different races because this kind of error can lead to a prison term for an innocent person.

SYSTEM VARIABLES

Postevent Information

In the 1970s E. F. Loftus began publishing research demonstrating that eyewitness memory reports can be affected by postevent information, or information that the witness encounters after the target event. E. F. Loftus showed that, if the postevent information is misleading, witnesses may subsequently report details that are consistent with that information instead of correct details. This phenomenon is known as the *misinformation effect*.

For example, consider an early experiment conducted by E. F. Loftus, Miller, and Burns (1978) in which witnesses viewed a slide sequence depicting a traffic accident. In one condition, the critical slide showed a car stopped in front of a yield sign at an intersection. Afterward the witnesses completed a questionnaire about the slides. The actual purpose of the questionnaire was not to test witnesses' memory but to manipulate the independent variable, which was the presence of misleading information. One version of the questionnaire contained the critical item "Did another car pass the red Datsun while it was stopped at the stop sign?" This item is misleading because it incorrectly identifies the type of sign. The control version of the questionnaire included the same item with the words "yield sign" replacing "stop sign."

After a 20-minute filler activity, the witnesses tried to identify the critical slide they had seen in a two-alternative recognition test. As would be expected, most (75 percent) of the control witnesses chose the correct slide. However, most (59 percent) of the misled witnesses chose the slide that was consistent with the misinformation they read instead of the slide they actually viewed.

In this experiment witnesses erred by substituting a suggested detail for a correct detail. E. F. Loftus also showed in other studies that witnesses can also be induced to report seeing nonexistent objects. For example, witnesses who had watched a video of an automobile accident answered a series of questions about it. Some witnesses were asked, "How fast was the white sports car going when it passed the barn while traveling along the country road?" This item is misleading because no barn appeared in the video. The control version of the question omitted the phrase "when it passed the barn." After a delay of one week, all witnesses completed a memory test that included the question "Did you see a barn?" Although most witnesses correctly answered "no," 17 percent of the misled witnesses reported seeing a barn, compared to only 3 percent of the controls.

The misinformation effect has been replicated many times by many researchers. The misleading information may be embedded in questions, narrative accounts, or pictures presented to the witnesses. Various kinds of target events have been used, as well as delay intervals of differing lengths. Across these studies, the typical result is that a substantial proportion of misled witnesses provide reports that are consistent with the misinformation they encountered. Sometimes witnesses will even provide descriptive details about nonexistent objects whose presence was suggested to them.

Researchers have proposed several hypotheses about the cause of the misinformation effect. In her early work, E. F. Loftus suggested that the misleading information might alter or replace the witnesses' memory trace of the target event so that the original trace no longer existed. This "overwriting" hypothesis is probably too simplistic, however, because it turns out that in some situations misled witnesses can retrieve the original, correct information. Partly in response to this finding, some researchers have argued that the misinformation effect is not the result of memory impairment but instead reflects strategic choices made by misled witnesses. For example, it is possible that witnesses can remember the target event perfectly well, but they realize that the experimenter expects them to report details consistent with the misinformation. Therefore, in an effort to be cooperative, they report what the experimenter wants to hear instead of what they actually remember. Besides this type of demand effect, another way that a strategic effect could emerge is if misled witnesses cannot recall the original details but assume that the misinformation must be accurate because it was given to them by the experimenter. This result is termed "misinformation acceptance." Witnesses' strategic choices probably account for some instances of the misinformation effect, but they cannot explain some data. For example, in several studies misled witnesses were asked during the memory test phase of the experiment *not* to report any details from the target event. Often these witnesses avoided reporting the misinformation, which implies that they truly believed that these details were part of the target event.

Two other explanations have some empirical support. First, the retrieval-blocking hypothesis says that, at the time of the test, memory traces for both the target event and the misinformation exist and are in competition with each other. If the misinformation trace is stronger (for example, because those details were encountered more recently), then that trace will not only be more accessible but will also block access to the accurate trace.

A final explanation is based on source monitoring. Many studies have demonstrated that people sometimes forget the source of information they have encountered (or they failed to encode the source in the first place). Consequently, they may misattribute the source. In other words, witnesses might remember the misinformation but wrongly believe those details were part of the target event rather than the postevent questionnaire. Consistent with this explanation, some data indicate that asking witnesses to pay special attention to the sources of the information they receive reduces suggestibility.

False Memories and the Effects of Imagining

Building on research on the effects of misleading information, researchers have found that it is possible to implant in witnesses false memories for events that never happened. For example, E. F. Loftus and Pickrell (1995) asked college students to read descriptions of events written by a relative (e.g., a parent or an older sibling). Some of the described episodes really happened to the participants, but one in each set was a false event about getting lost in a shopping mall as a small child, and how the participant cried before being found by an elderly woman and reunited with relieved family members. The participants initially wrote down any memories they had of the events. Then, they were interviewed twice during a four-week period, each time trying to recall as much as possible about the events while using their relatives' descriptions as cues. This procedure induced 25 percent of the participants to say during the final interview that they remembered the false event (by comparison, 68 percent of the true events were remembered), although they rated the false event as less clear than the true ones, and they were less confident about those memories.

It appears that asking participants to imagine an event is a fairly effective method of creating a false memory. Garry, Manning, E. F. Loftus, and Sherman (1996) found that instructing participants to visualize a childhood event and guiding them as they imagined different aspects of it significantly increased their ratings of the likelihood that the event occurred. Other studies suggest that people who are skilled at forming clear mental images may be especially susceptible to creating false memories, and imagining can increase people's belief that they can *remember* experiencing an event, including contextual details, instead of simply becoming more convinced that they *know* the event occurred. Researchers have proposed that imagining leads to the creation of plausible and vivid details associated with the event, and later participants may become confused about the source of those details, believing they were part of the actual event.

A potential criticism of studies of false memories is that perhaps the researchers have assumed that the "false" event did not occur in a participant's life, but actually it did. Studies designed to address this issue have asked participants about events that could not possibly have happened, such as medical procedures that never occur in the participants' home country, with the same results as previous studies.

Because of the abundance of data showing how false memories can be created, researchers have cautioned against the use of imagining as a tool to help retrieve memories. For example, police investigators have sometimes asked witnesses to imagine in an effort to assist them in remembering the details of a crime, and therapists have used guided imagery to help clients remember and deal with traumatic childhood memories. However, the data indicate that such techniques are likely to be counterproductive and could even lead people to "remember" episodes they never experienced.

Lineups

After police investigators have identified a suspect, they may ask witnesses to view a photo lineup (or, less commonly, a live lineup) in an effort to determine whether the suspect is the perpetrator. Researchers have examined the relationship between a number of variables and identification accuracy. Some of those variables are reviewed below.

The Selection of Foils

A lineup includes one person whom the police suspect of committing the crime being investigated. This person may or may not be guilty. The other members of the lineup are not the focus of the investigation, and they are called fillers or foils, as noted previously. Researchers have discovered that the physical appearance of the foils can influence a witness's choice. For example, if the suspect is innocent but resembles the perpetrator significantly more than the foils do, then the witness may mistakenly identify that suspect as the guilty party (a false alarm). Although selecting foils who look like the suspect may at first seem to be the fix for this problem, this strategy may lower the hit rate for target-present lineups because the lineup members may be so similar that they are indistinguishable. Studies show that it is actually better to use foils who match the witness's description of the perpetrator.

Instructions

Witnesses sometimes assume that a suspect arrested by the police is probably guilty and therefore that a lineup probably includes the perpetrator. They may also feel intense pressure to choose someone from a lineup in order to help solve the crime. Researchers studying police transcripts have found instances in which lineup administrators encouraged witnesses who had initially failed to identify anyone to take another look or pressured witnesses by telling them that they had to help get a dangerous criminal off the streets. However, even without explicit pressure from the administrator, witnesses may feel compelled to choose. To counteract this tendency, administrators can simply inform witnesses that the perpetrator may or may not be in the lineup. Studies show that this instruction significantly reduces false alarms in target-absent lineups without harming hit rates in target-present lineups.

Confidence

After a witness makes an identification (or, alternatively, says that the perpetrator is not in the lineup), the police investigator, in an effort to determine whether the witness was accurate, may ask him or her for a rating or estimate of confidence. It seems reasonable to assume that highly confident witnesses are probably accurate and uncertain witnesses are probably wrong. Not surprisingly, jurors are profoundly influenced by a witness's expression of confidence as they assess credibility during a trial. However, the strength of the actual correlation is affected by several variables such as whether the witness had a good opportunity to view the perpetrator and the degree of resemblance between the perpetrator and the innocent suspect (in the case of false alarms). Moreover, these variables can be difficult to estimate. As a result, from a practical viewpoint, the usefulness of confidence as a predictor of accuracy is limited.

A further problem with confidence is that it is malleable, meaning that witnesses' confidence can be changed by exposure to information that may be unrelated to their performance in the lineup task. For example, Luus and Wells (1994) had pairs of witnesses watch a staged theft. After each witness independently completed a lineup task, the researchers told him or her that the other witness had either identified the same person, identified another person, or decided that the target was not present (there was also a control condition in which no information about the co-witness's choice was provided). Finally, the witnesses rated their confidence. Telling witnesses that the co-witness had identified the same person caused an increase in confidence relative to the control condition, whereas telling them that the co-witness identified someone else or no one caused a drop in confidence. Interestingly, almost all of the witnesses in this study made a mistaken identification due to the procedures the researchers used. Luus and Wells argued that their findings could easily apply to real-life investigations because lineup administrators might tell one witness what another said, or the witnesses might exchange information directly if they come into contact with one another.

Another way to manipulate confidence is to give witnesses feedback on their own performance. It is not illegal for police investigators to give witnesses confirming feedback (e.g., telling them that they did a good job and identified the actual suspect) or disconfirming feedback (e.g., telling them that they chose someone who was not the focus of the investigation). However, this information might be incorrect because the police could have targeted an innocent person. Wells and Bradfield (1998) asked participants to watch a surveillance video that showed a man who later shot a security guard in a retail store. Afterward, the witnesses viewed a lineup and then were given randomly selected feedback. The authors found that confirming feedback inflated confidence relative to disconfirming feedback. Witnesses who received confirming feedback also said that they had a clearer view of the perpetrator, that they could more clearly see the details of his face, that they paid closer attention to the event, and that it was easier for them to make an identification. These witnesses, like those in Luus and Wells's study, identified someone other than the actual perpetrator. However, jurors would no doubt find them quite convincing.

To reduce the likelihood that the lineup administrator will give feedback to the witness (either verbally or through nonverbal signals such as body language), researchers advise that a double-blind procedure should be used. In other words, the administrator should not be the police detective who is conducting the investigation of the case, but instead should be someone who does not know which lineup member is the suspect and which members are foils. Alternatively, if the lineup involves photos rather than live members, it could be presented by a computer. If one of these methods is used, there is less chance that the administrator could intentionally or inadvertently influence the witness's confidence.

Simultaneous Versus Sequential Lineups

In the United States, most lineups include six members, or sometimes five. Traditionally, police investigators have presented all the members at once to witnesses, asking them to look at the entire lineup and decide whether the perpetrator is in it. However, researchers recommend using a sequential lineup instead. In the sequential procedure, witnesses are told that they will view an unspecified number of individuals one at a time. When a particular lineup member is shown, the witness must decide whether that person is the perpetrator. If the decision is "yes," the procedure is finished. If the decision is "no," the administrator moves on to the next lineup member. The process continues until the witness either identifies someone as the perpetrator or views all the members without identifying anyone.

The problem with the simultaneous lineup is that it encourages relative judgments. That is, witnesses compare the members and choose the one who most resembles the perpetrator. This strategy can lead to errors because there will always be one member who looks at least a little bit more like the perpetrator than anyone else, even if the resemblance is not strong. Therefore, if the lineup does not contain the guilty person, a false alarm is likely.

It is difficult for witnesses to make relative judgments when sequential lineups are used because the six members cannot be viewed at the same time and compared. As a result, witnesses tend to make absolute judgments, separately comparing each member to their memory of the perpetrator, and the number of false alarms decreases. A potential problem with sequential lineups is that they appear to reduce the number of hits as well as false alarms; in other words, these lineups induce witnesses to make fewer identifications, whether accurate or not. However, some researchers have argued that the sequential lineup merely weeds out lucky guesses without affecting the hit rate for witnesses who truly remember the perpetrator and recognize him or her in the lineup. In any case, researchers typically advocate the use of sequential lineups, largely to cut down on the chance of convicting innocent individuals.

SUMMARY

It is clear that eyewitnesses can make mistakes, even if they are trying to be honest and cooperative. Their reports or lineup identifications can be distorted by stress, the presence of a weapon, exposure to misleading information, or suggestive questioning methods that encourage them to imagine what happened. They are less able to identify a perpetrator accurately if that person is of a different race. Several additional variables can affect their lineup performance or the confidence they express about their choice.

The results of the many eyewitness memory studies that have been conducted since the 1970s have allowed researchers to formulate a number of recommendations for police investigators and others who work within the legal system. Use of these recommendations could lead to better techniques for questioning witnesses and administering lineups, as well as a better understanding of the variables that limit eyewitness memory performance. Drawing upon empirical findings, researchers have developed procedures such as the cognitive interview (CI) to help police interviewers get as many accurate details as possible from witnesses while minimizing the amount of incorrect information reported. The CI includes several steps that the traditional police interview lacks. For example, the CI instructs interviewers to build rapport with witnesses as a way of putting them at ease and to let them control the pace and direction of the interview while refraining from interrupting them. Interviewers should also ask witnesses to "reinstate the context" that was present during the target event by recalling the physical environment (e.g., objects, smells, or sounds) and their emotions and thoughts at the time. These contextual details can serve as cues that help witnesses retrieve event details from memory.

Before the 1990s law enforcement personnel, prosecutors, and judges showed little interest in the eyewitness literature and in researchers' suggestions. However, the situation changed when DNA testing began to exonerate innocent persons who had been convicted of crimes and the press began to publicize these cases. Realizing that there was room for improvement within the system, courts and police departments have become more willing to establish reforms based on scientific findings. In fact, in the late 1990s, Attorney General Janet Reno ordered the National Institute of Justice to oversee a committee of researchers, attorneys, and police officers who worked together to develop a set of national guidelines for interacting with eyewitnesses. A few U.S. jurisdictions have adopted these guidelines. In the future, perhaps even more will join in the effort to improve the performance of eyewitnesses and make sure that the evidence collected from them is used appropriately.

REFERENCES AND FURTHER READINGS

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based model. *Psychonomic Bulletin and Review, 5,* 1–21.

Christianson, S.-Å., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Emotion and Cognition, 5,* 81–108.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (1994). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28,* 687–706.

Easterbrook, J. A. (1959). The effect of emotion on cue-utilization and the organization of behavior. *Psychological Review, 66,* 183–201.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin and Review, 3,* 208–214.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7,* 560–572.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus." *Law and Human Behavior, 11,* 55–62.

Loftus, E. F., Miller, H. J., & Burns, D. G. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19–31.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25,* 720–725.

Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79,* 714–723.

Mazzoni, G., & Memon, A. (2003). Imagination can create false autobiographical memories. *Psychological Science, 14,* 186–188.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7,* 3–35.

Mitchell, K. J., Livosky, M., & Mather, M. (1998). The weapon focus effect revisited: The role of novelty. *Legal and Criminological Psychology, 3,* 287–303.

Peters, D. P. (1988). Eyewitness memory and arousal in a natural setting. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 89–94). Chichester, UK: Wiley.

Peters, D. P. (1991). The influence of stress and arousal on the child witness. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 60–76). Washington, DC: American Psychological Association.

Peters, D. P. (1997). Stress, arousal, and children's eyewitness memory. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events* (pp. 351–370). Mahwah, NJ: Erlbaum.

Pickel, K. L. (1999). The influence of context on the "weapon focus" effect. *Law and Human Behavior, 23,* 299–311.

Pickel, K. L., Ross, S. J., & Truelove, R. S. (2006). Do weapons automatically capture attention? *Applied Cognitive Psychology, 20,* 871–893.

Reisberg, D., & Heuer, F. (in press). The influence of emotion on memory in forensic settings. In M. P. Toglia, J. D. Read, J. D. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology: Vol. 1. Memory for events* (pp. 81–116). Mahwah, NJ: Erlbaum.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83,* 360–376.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest, 7,* 45–75.

48

Repressed and Recovered Memory

Beverly R. King

University of North Carolina at Pembroke

In 1885 the first person to experimentally investigate the properties of memory, German psychologist Herman Ebbinghaus (1885/1913), published information on the curve of forgetting, which showed the progression of forgetting with the passage of time. Essentially, this curve demonstrated that (up to a point) as more time elapses, more information is forgotten—at least in the case of rather innocuous material such as nonsense syllables. Ebbinghaus's observations are still valid, more than 120 years later. In 1932 British psychologist Sir Frederic C. Bartlett concluded that human memory is a reconstructive rather than a reproductive process. That is, our memories are never exact duplicates of the experiences they purport to capture but rather are rebuilt from bits and fragments with each recollection. Both cognitive psychologists and neuroscientists have accumulated research evidence that Bartlett's observation was accurate (Schacter, 1996). Memories typically are less available over time and may not be recalled with absolute accuracy, but is it possible for a person to have complete amnesia for an entire experience and then retrieve relatively intact memories of that experience after a lengthy period of time has elapsed? This is the question at the heart of one of the most contentious debates in the history of psychology—the recovered memory debate.

Although psychologists had discussed the notion of memory repression at least since Freud, it garnered special attention in the late 1980s and early 1990s with the publication of books such as *The Courage to Heal* (Bass & Davis, 1988). These books encouraged individuals who knew that they had been sexually abused to seek help and encouraged others to think about whether they might have been sexually abused based on the possession of particular symptoms sometimes presented in the form of a checklist. The implication was that memories such as those of childhood sexual abuse (CSA) might be put completely out of awareness (or "repressed") due to their traumatic nature but could be subsequently recalled (or "recovered") with the help of certain techniques and/or an experienced therapist.

In a number of cases, adult women came forward to report memories of CSA recovered in therapy but later recanted their stories (Loftus, 1997) after therapy was discontinued. In addition to the stress undoubtedly engendered by the recovery of "false" memories, some of these women accused family members of the abuse and even brought criminal charges against their alleged abusers before recanting (Ost, 2003). Thus began what Ost and others have referred to as "the memory wars." The "wars" center on whether it is actually possible to repress and later recover memories of trauma. The two sides of the debate are summarized as follows: On one hand are those, often mental health practitioners, who believe that traumatic events such as CSA may be stored by a different process than "normal" memories such that they are difficult to recall but can be retrieved relatively intact years later with the right cues. This special storage process or amnesia protects victims of CSA from the pain of these traumatic memories. On the other hand are individuals, often academics or research scientists, who claim that there is no empirical evidence for this special process (American Psychological Association [APA], 1995; Clancy, Schacter, McNally, & Pitman, 2000) and that traumatic events typically are not forgotten, although people may try not to think about them for long periods of time (McNally, 2004). They also point out that in many incidences of so-called recovered memories, the memories were "recalled" after particular media exposure (e.g., television shows or self-help books about repressed memories) or following participation in inappropriate, suggestive, and perhaps unethical therapy techniques.

No one, on either side of the debate, disagrees that sexual abuse of children occurs or that forgetting can occur. In both retrospective and prospective studies, researchers have found some adults who had periods in their lives during which they could not recall a traumatic event such as sexual or physical abuse (see Berliner, Hyman, Thomas, & Fitzgerald, 2003, for a review of some of these studies). Some psychiatric disorders even have as part of their diagnostic criteria amnesia for all or part of a traumatic episode (e.g., post traumatic stress disorder [PTSD] and acute stress disorder; American Psychiatric Association, 2000). However, there usually is consensus that most individuals who were sexually abused as children remember all or part of what happened to them although they may not fully understand or disclose it (APA, 1995). An extensive literature based on research with both animals and humans indicates that traumatic memories, especially the central details, are not easily forgotten (Alison, Kebbell, & Lewis, 2006).

The question remains: is it at all possible for a person to have complete amnesia for an entire experience and then retrieve relatively intact memories of that experience after a lengthy period of time has elapsed? Due to ethical constraints, we cannot test this issue directly, and neither clinicians nor researchers would argue that it is impossible for a memory to be repressed and later recovered. After all, as McNally (2004) has pointed out in regard to this issue, "one can never prove the null hypothesis that a phenomenon does not occur" (p. 100). However, several lines of research over the past two decades have focused on the following questions:

1. *Memory for actual traumatic events:* How likely is it that memories of traumatic events will be recalled, both shortly after the events have occurred and years later?
2. *Memory distortion:* How, and under what conditions, might recollections of past events become distorted?
3. *False memory creation:* Is it possible for individuals to construct entirely false memories (or "pseudomemories") for events that never occurred but in which they strongly believe? Especially of interest has been whether or not techniques employed in psychotherapy can lead to the creation of false memories.

THEORIES

The concept of repression comes from Freudian theory. To Freud (1910), the process of repression involved driving a memory out of the conscious into the unconscious mind even though the repressed material could continue to have influences on behavior. Repression could occur either during or after an event and was one mechanism that could lead to amnesia for traumatic events (Allison et al., 2006). Dissociation was a different mechanism, in which memories of traumatic events were somehow disrupted and which resulted in memory impairment rather than full-blown amnesia (Allison et al., 2006). Currently, theorists use the term "dissociative amnesia" to describe a *condition* of memory loss greater than ordinary forgetfulness (Porter, Campbell, Birt, & Woodworth, 2003) and "repression" to describe one *process* by which an event might be forgotten, at least for a while.

Two other theoretical models are of relevance to the repressed and recovered memory issue. Both of these would relate to how false memories might be created. One is a model such as that of Smith (2000), which describes how memories are constructed upon recollection. In Smith's model,

> The raw material of memory is a set of stored fragments… Memory retrieval is an iterative process involving interactions between memory fragments and a coherent jigsaw, called the *Current Memory,* which is assembled from the memory fragments.…In assembling this jigsaw, a measure of goodness-of-fit is computed (*Harmony*) and this corresponds to the confidence the memory retriever might feel about the authenticity of the retrieved information. (p. 246)

In other words, just as Bartlett observed in 1932, memory is a reconstructive process. When bits and pieces of memory fragments are put together for recollection, the memory retriever might have a great deal of confidence that the memory is being recalled accurately when, in fact, it may be that the "measure of goodness-of-fit" is computed incorrectly due to other forces such as emotions, imagination, or familiarity.

The second relevant model is one that describes how false memories might be created (Loftus & Bernstein, 2005). In this model (Mazzoni, Loftus, & Kirsch, 2001), a three-step process leads to the formation of false memories:

1. Individuals must see events to be incorporated into memory as plausible.
2. They must come to believe that they had experienced the event.
3. They must reinterpret narratives or images about events to produce what they feel are genuine memories.

A large body of empirical evidence exists that indicates not only that the creation of false memories is possible but also that it is surprisingly easy to implant them in the minds of research participants.

METHODS AND RESEARCH RESULTS

Support for the Mazzoni et al. (2001) model and research related to false memory creation is discussed following presentation of research methods and findings relevant to memory for actual traumatic events and memory distortion.

Memory for Actual Traumatic Events

Are memories for traumatic events remembered more vividly than memories for everyday events (the trauma superiority argument), or are memories of traumatic events somehow encoded, stored, or retrieved by different processes such that they can be forgotten for long periods of time (the traumatic memory argument; Alison et al., 2006)? This question is typically addressed by asking individuals who have experienced past trauma to recount their experiences, either directly or while being questioned on another issue. Overwhelmingly, the research supports the trauma superiority argument—memories for stressful experiences are not easily forgotten, especially the central details of the events. For example, in studies of children who have experienced traumatic events such as parental homicide, physical abuse, or holocaust experiences, they remember these experiences very well and can provide detailed, vivid accounts of the events (see McNally, 2003, for a review). Similarly, in both retrospective and prospective studies with adult respondents, traumatic memories for events such as genocide, torture, disasters, and accidents are usually remembered both continuously and vividly, with central details more likely to be remembered than peripheral details (see McNally, 2003, and Porter et al., 2003, for reviews). Certainly forgetting can occur, not only for relatively ordinary events but also for more significant ones (Porter et al., 2003). However, if adult victims of trauma have forgetfulness problems after their experiences, it is usually everyday forgetfulness and not an inability to remember the event(s) (McNally, 2003).

In studies specifically focused on participants who had experienced sexual abuse as children, a minority (percentages vary by study) do state that there was a period of time in their lives when they had no recall of the experience and recovered the memory after a period of forgetting. In a national probability sample of 711 women in the contiguous United States, more than a fourth of those who reported sexual coercion or abuse while growing up indicated that they had forgotten the abuse for some length of time and later remembered it on their own (S. C. Wilsnack, Wonderlich, Kristjanson, Vogeltanz-Hohn, & R. W. Wilsnack, 2002). However, in the majority of historical cases of abuse, individuals bringing the complaints report that they had always been able to recall the incident(s) since its (their) occurrence but simply chose to delay disclosure (Porter et al., 2003). The APA (Alpert, 1996) maintains that most victims of CSA remember at least some of what happened to them continuously from the time of abuse.

There are a number of reasons that individuals may not think about a traumatic experience such as CSA for an extended period of time. For example, they may not have interpreted the experience as traumatic when it happened (they had no schema for the experience in childhood); they may have intentionally suppressed the experience; they may simply choose not to talk about abusive experiences for various reasons; or they may have truly forgotten the experience (McNally, 2003; Porter et al., 2003).

Goodman and colleagues (2003) contacted 170 individuals who were victims of CSA and who had participated in a study of court testimony 13 years earlier. They surveyed these individuals three times: first by phone, then by questionnaire, and finally in a face-to-face interview. Only slightly over 8 percent of these individuals did not report the abuse in any phase of the study, and they were more likely than those who did report to have been under age five at the time of the abuse and to have been victims of less severe abuse.

Memory Distortion

One of the leading researchers on issues related to repressed and recovered memory is Elizabeth Loftus. Her work on how memory can be distorted began in the early 1970s when she began studying how leading questions can influence memory; later, she studied what has come to be called "the misinformation effect" (Loftus & Hoffman, 1989). Literally hundreds of studies have now been conducted by Loftus and her students, all of which show that when people witness an event and are later exposed to new and misleading information about it, they often "misrecall." Their recollections become distorted as the result of suggestion (Loftus, 1997). In the real world, this information might come from media sources or interrogators; in the laboratory, the information is fed to participants by the researcher.

In one of her earliest studies, Loftus showed research participants films of traffic accidents and then asked questions in various forms about what they saw (Loftus & Palmer, 1974). Individuals who were asked "How fast were the cars going when they *smashed* into each other?" gave faster speed estimates than individuals who were asked "How fast were the cars going when they *hit* each other?" Using the word *smashed* versus the word *hit* also led more individuals to report seeing broken glass when there was no broken glass in the original film. Leading questions can distort memory, as can information coming from a source other than the event to be remembered. In another study, participants saw a simulated automobile accident; half of the participants then received misleading information about the accident (the false suggestion that the stop sign they saw in the simulated accident was actually a yield sign). These individuals were much more likely to have inaccurate memories (to incorporate the yield sign into their memories) than the research participants who did not receive misleading information (Loftus, Miller, & Burns, 1978).

In addition to recalling yield signs instead of stop signs, people in these types of studies have recalled nonexistent tape recorders, a blue vehicle as white, and Mickey Mouse as Minnie (see Loftus, 2005a, for a review of Loftus's 30 years of investigations on this topic). Hundreds of studies on the misinformation effect show that misleading information "…can change an individual's recollection in predictable, and sometimes very powerful, ways" (Loftus, 2005b, p. 1).

False Memory Creation

Psychological research has demonstrated not only that memory can be distorted but also that entirely false whole memories can be created. These rich false memories involve "the subjective feeling that one is experiencing a genuine recollection, replete with sensory details, and even expressed with confidence and emotion, even though the event never happened" (Loftus & Bernstein, 2005, p. 101). Loftus and Pickrell (1995) conducted the first research study in which an attempt was made to create an entire memory for something that never happened. The procedure they used to do this has been called the *familiar informant false-narrative procedure* (Lindsay, Hagen, Read, Wade, & Garry, 2004) or, more simply, the *lost-in-the-mall* technique (Loftus, 2003b). In this study, participants read short narratives—all of which they thought were true stories supplied by family members; however, one was a false event. The false event was a plausible shopping trip to a mall during which the participant was ostensibly lost for an extended period of time but eventually rescued and comforted by an elderly woman and returned to the family. After participants read each story, the researchers instructed them to write what they remembered about the events described. In two follow-up interviews, participants provided as much detail about the memories as they could and described how their memories compared with those of the relative who provided the information. Parts of the narratives were provided as retrieval cues. A majority (68 percent) of participants recalled something about the true events; 29 percent remembered, either completely or in part, the false event after the initial reading; and 25 percent continued to assert that they remembered the false event in the follow-up interviews.

The Loftus and Pickrell (1995) study was followed by numerous studies that showed that both strong and relatively subtle suggestions could lead to people believing "that they had experiences in childhood that they almost certainly did not have" (Loftus, 2003a, p. 108). For example, in several studies by Hyman, Husband, and Billings (1995), college students were asked to recall events from their childhood. Some of the events were true events as reported by parents and some were false events, such as a hospitalization, a birthday party, spilling punch at a wedding, or evacuating a grocery store when an overhead sprinkler system went awry. No student recalled the false event in a first interview, but in a second interview 18 to 20 percent recalled something about the false event. A long list of subsequent studies further documents the types of false memories that research participants can be lead to believe: being attacked by an animal (Porter, Yuille, & Lehman, 1999); falling off a bicycle and receiving stitches in their leg (Heaps & Nash, 1999); and being saved from drowning (Heaps & Nash, 2001), to name a few.

These types of studies have been criticized on two fronts:

1. Perhaps instead of implanted false memories, participants were simply recalling memories of events that had really happened to them. Maybe several participants in the lost-in-the-mall study really had been lost at some point in childhood and were simply elaborating on an existing memory.
2. More pertinent to the repressed/recovered memory debate, all of these events were relatively common or at least entirely plausible; that is, they could have happened. Wouldn't it be much less likely that someone could, under suggestion, recall a traumatic event such as sexual abuse that never actually happened?

Researchers have devised several techniques to address both of these criticisms. In each technique, researchers have made an attempt to implant memories for events that could not have happened or that are highly improbable. Pezdek and colleagues (e.g., Pezdek, Finger, & Hodge, 1997; Pezdek & Hodge, 1999) found that less plausible events (e.g., having had a rectal enema as a child) were less successfully implanted than more plausible events. However, some researchers have been successful at implanting quite implausible memories. Mazzoni et al. (2001) used the not just implausible but bizarre target event of seeing, as a child, someone possessed by demons. All participants in this study rated a list of events in terms of the plausibility that they could have happened to other people and the confidence that they had not been experienced by the participants. Then, some of the participants read fake articles that described demonic possession as common; these same participants then took a fear survey and had it interpreted for them, indicating evidence that they had witnessed demonic possession in childhood. Finally, all participants repeated the plausibility and confidence ratings. Participants who had been exposed to the fake articles and survey results thought it was more plausible and more likely that they had indeed witnessed demonic possession as a child. The manipulation was effective even with people who entered the study believing that demonic possession was rather implausible.

This study by Mazzoni et al. (2001) demonstrates the stages in the process by which false memories may be created. Research participants came to see demonic possession as *plausible* though the use of fake articles that portrayed demonic possession as common and came to *believe* that they had experienced the event by virtue of survey results that they thought indicated a history of witnessing demonic possession. What might lead people to come to believe that they have experienced an event even when they have not? Other studies indicate that belief may be created through a process of "imagination inflation" (imaging an event can increase confidence that one has actually experienced the event; Garry, Manning, Loftus, & Sherman, 1996). Garry et al. (1996) asked research participants to indicate the likelihood that each of 40 events had happened to them during childhood. Two weeks later, they were instructed to imagine that some of these events had happened to them (e.g., breaking a window with their hand) and, later again, were asked again to indicate the likelihood that each of the 40 events had happened to them. For the broken window scenario, 24 percent of those who imagined the event reported increased confidence compared to 12 percent of those not asked to imagine the event.

Thomas and Loftus (2002) investigated imagination inflation for more unusual experiences. Participants were asked to perform or imagine some ordinary actions and some more unusual ones (e.g., sitting on dice). Later, they imagined some of the actions zero to five times. Imagination inflation occurred even for the unusual actions; the more participants imagined an action, the more inflation occurred. Why might the act of imagining an event increase a person's belief that the event happened to them in the past? Loftus (2005a) posits two possible reasons: (a) imagining an event might remind some participants of a true experience from their past or, more likely, (b) imagining made an event seem more familiar to a participant, and the familiarity was mistaken as childhood memories.

Another interesting set of studies further illustrates how repeated exposure can raise the likelihood of false memory formation perhaps by increasing familiarity. In these studies, researchers asked participants to evaluate advertising copy for a Disney resort. Some of the participants saw a fake Disney ad containing a picture of Bugs Bunny and later were asked about any trips to Disney they had taken as children. Of individuals in one study (Braun, Ellis, & Loftus, 2002), 16 percent who were exposed to the fake ad claimed to have met Bugs Bunny when they visited Disney (an impossibility, as Bugs Bunny is a Warner Bros. cartoon character). Repeated exposure to the fake ad raised the false memory rate to 25 percent in one study and 36 percent in another (Grinley, 2002). In one of these studies, participants who falsely recalled having met Bugs at Disney had quite detailed recollections; some remembered shaking his hand; others recalled hugging him; and still others had recollections of touching his ear, touching his tail, or hearing him say his familiar line, "What's up, Doc?"

False memory rates can be boosted even higher when photographs are used as memory cues. Wade, Garry, Read, and Lindsay (2002) gave research participants photos of themselves as young children and asked them to recall the events depicted in the photos. Most of the photos were genuine but one for each participant was faked by digitally altering a photo of a hot-air balloon ride to include a real childhood image. Of participants, 50 percent recalled some memory of the hot-air balloon and often provided detailed descriptions of the ride despite the assurance of family informants that this event had never happened. In response to the criticism that studies using "doctored" photographs lack ecological validity, Lindsay, Hagen, Read, Wade, and Garry (2004) used a real photograph to help implant a false memory. Over two interviews, they asked participants to remember three school experiences—two were real but the earliest was false. The false event involved putting toy slime in a teacher's desk. All participants received a narrative of the experience; half also received a class picture. Less than half of the narrative-only participants reported slime images or images with memories; about two-thirds of the photo participants did. The authors speculated that three different mechanisms might explain the high incidence of false memories in their study: (a) The photo may have added additional authoritativeness to the narrative; (b) the photo may have lead participants to speculate about details related to the false event; and (c) memories of details in the photo may have been mixed with imagination, leading to the creation of vivid images of the false event (a source monitoring error—see Comparisons section).

APPLICATIONS

Given that the most heated arguments regarding whether memories can be repressed and then recovered are in cases of CSA, the memory research described is most applicable to judges, juries, and legal experts, who must assess the credibility of recovered memories of CSA, and to therapists, who might inadvertently lead their clients to develop false memories of CSA. Both Alison et al. (2006) and Porter et al. (2003) present recommendations and guidelines for judges, juries, and experts to help in evaluating allegations of CSA based on recovered memories. Neither set of authors argues that every case of alleged recovered memory of CSA is false; no doubt some cases of alleged historical abuse are accurate. How can the criminal justice system distinguish between the two? There is often no corroborating evidence in alleged cases of historical abuse, and there is always the fear of false convictions. Alison et al. (2006) suggest that each case should be evaluated on its own merits alongside research-based information. These authors advocate an approach in which supporting evidence is required to establish credibility in recovered memory cases.

Porter et al. (2003) provide four guidelines for deciding on the veracity of historical allegations that judges, juries, expert witnesses, lawyers, and police investigators could use. The first guideline deals with the context in which memories are recalled. When recovery of memories is accomplished by use of suggestive techniques, there is a greater likelihood of illusory memories being recovered than if memories are recalled spontaneously. The second and third guidelines deal with differences in the content of recalled memories or among persons who recover mistaken memories (both of which are described more fully in the Comparison section). Although more research is needed, researchers have found certain content differences between real and mistaken memories. Individuals with a greater susceptibility to suggestion and a tendency toward dissociation (and perhaps those who are more introverted and imaginative) tend to recall more mistaken memories than those who are low in these characteristics. Finally, in line with the recommendations by Alison et al. (2006), Porter et al. (2003) suggest that corroboration of any alleged event involving historical CSA should be sought. Corroboration adds credibility to a claim; lack of it raises doubts.

Probably more than any other application, attention has been directed toward whether therapeutic techniques can foster false memory recollections and what advice should be given to therapists related to this issue. Recovered memories that emerge in the process of psychotherapy have engendered one of the most heated aspects of the repressed memory controversy, not only because of the difficulty in distinguishing true from created memories but also because of ethical considerations—the lives of clients and their families can be significantly altered if CSA is suspected where none actually existed.

Researchers have applied the term “memory work” to psychotherapy techniques used to help individuals retrieve ostensibly repressed memories of childhood sexual abuse (Loftus, 1993). Some of the techniques in this category are guided imagery, suggesting false memories, hypnosis, searching for early memories, age regression, hypnotic age regression, past-life regression, symptom interpretation, bogus personality interpretation, dream interpretation, physical symptom interpretation, and bibliotherapy (Lynn, Lock, Loftus, Krackow, & Lilienfeld, 2003). How might psychotherapeutic techniques encourage the development of false memories? First of all, demand characteristics may be operating in a therapeutic setting, just as they frequently do in research settings. In research, *demand characteristics* refer to the cues that may influence the behavior of participants, perhaps by suggesting an experimental hypothesis. In psychotherapy, the term may refer to the cues that convey the therapist’s expectations to clients and influence their behavior (Kanter, Kohlenberg, & Loftus, 2002). A client may feel he or she needs to produce memories consistent with suggestions from a therapist simply because the therapist is seen as a trusted authority figure (Alison et al., 2006). In fact, the effects of suggestive techniques may be more pronounced in a clinical than in a laboratory setting because the influence of a regularly seen and trusted therapist on a vulnerable patient may be much greater than the influence of an experimenter on a research participant seen once or only a few times and who is participating for money or course credit (Lynn et al., 2003).

Secondly, all the research discussed above indicates that these techniques can not only help a person retrieve memories but also help create memories. (See Lynn et al., 2003, for a full review of these techniques and how they may be used to help create memories.) Loftus and Pickrell’s (1995) lost-in-the-mall studies as well as others using the same methodology have demonstrated that, when scenarios are suggested to individuals and they are asked to either imagine or otherwise think about the events, they may create detailed memories that are obviously false but which they come to believe are true parts of their history. These suggestions may come in a variety of forms such as symptom interpretation or guided imagery (Lynn et al., 2003). Checklists of symptoms that therapists or authors may suggest are indicative of a history of abuse (but which in reality could apply to almost anyone) may lead some people to create a history to match the symptoms. Even more likely to lead to the formation of false memories would be a therapist’s suggestion that a person has certain personality characteristics indicative of abuse. The APA has been clear in its assertion that there is no one set of symptoms that indicates that someone has been abused. A therapist or book author who suggests otherwise is not basing his or her assertion on empirical evidence (APA, 1995).

Another psychotherapeutic technique that can be problematic in terms of leading to false memory creation is guided imagery. In guided imagery, clients “are instructed to relax, close their eyes, and imagine various scenarios described by the therapist” (Lynn et al., 2003, p. 207). Often used successfully in systematic desensitization (where imagery is combined with relaxation techniques to treat a phobia), the technique is more controversial when used to help a client try to recall repressed traumatic memories. The studies of imagination inflation discussed above show that imagining events makes them seem more familiar and increases confidence in their occurrence. This familiarity and confidence can be misattributed to true childhood memories.

Another pertinent question is: How many clinicians actually use directive therapies designed to help recover repressed memories? Polusny and Follette (1996) surveyed a random sample of clinical and counseling psychologists and found that 25 percent of them reported using guided imagery, dream interpretation, bibliotherapy regarding sexual abuse, and free association of childhood memories with clients who had no specific memories of CSA. In another survey of British and American doctoral-level psychotherapists and counselors conducted by Poole, Lindsay, Memon, and Bull (1995), 25 percent of respondents thought that memory recovery was an important part of therapy and 71 percent of therapists reported using suggestive memory-recovery techniques (including hypnosis, dream interpretation, age regression, and guided imagery) when they suspected CSA. They also reported a wide variety of behavioral symptoms they thought could indicate child sexual abuse including sexual dysfunction, relationships problems, low self-esteem, and depression. However, 90 percent of these therapists were aware that clients could have mistaken memories; more recent surveys also indicate widespread awareness of the possibility of false beliefs concerning CSA (Gore-Felton et al., 2000). This finding, coupled with the report in a recent national probability sample of women in the United States that only 1.8 percent had recovered memories of sexual abuse with the help of a therapist (S. C. Wilsnack et al., 2002), indicates that the problem of false recovered memories in therapy may be less prevalent than in previous decades.

Researchers and theorists have written many articles with recommendations for therapists related to suspected CSA and potentially repressed memories. Some of the recommendations are as follows:

- Therapists should not suggest an explanation of abuse but rather stick with information reported by the client (APA, 1995).
- Clinicians should be aware of how much they can influence their clients' recollections and exercise restraint in using imagery to help recover assumed memories (Loftus, 1997).
- Mental health professionals should consider all alternative explanations for a recovered memory (e.g., that the person simply hasn't thought about the experience in a long time) and consider the plausibility of the memory (Taylor, 2004). For example, the average age of a first autobiographical memory is 3 ½ years; a memory of something that happened substantially earlier than this is highly unusual.

The APA (1995) also recommends that individuals seeking a therapist be cautious if it appears that a large number of the therapist's clients have recovered memories of childhood sexual abuse while in treatment.

COMPARISONS

Comparisons have been made between persons in terms of their propensity to distort memories and between recollections of true versus false memories. Research comparing individuals with suspected repressed memories with those with continuous memories and controls has been conducted by McNally and colleagues (e.g., Clancy, McNally, Schacter, Lenzenweger, & Pitman, 2002; Clancy et al., 2000) using variants of a research design called the Deese/Roediger-McDermott (DRM) paradigm. In this procedure, the researcher presents participants with lists comprising words related to a theme word; however, the theme word is not present in the list. For example, a list containing the words *sour, candy, sugar,* and *bitter* would not contain the primary theme word of *sugar.* After presenting the lists, the researcher gives the participants a recall test and then a recognition test (in which they select list words from among presented words, nonpresented theme words, and nonstudied words). From these tests, researchers identify false recall and false recognition (both of which involve individuals indicating that they studied a nonpresented theme word).

People who report recovered memories of highly unlikely events show a tendency for false memory formation in the DRM paradigm. Clancy et al. (2002) investigated memory distortion in people claiming to have recovered memories of abduction by space aliens. This group was compared to two other groups: (a) people who believed that they had been abducted but had no memories of the event (repressed group); and (b) people who said they had no history of abduction by aliens (control group). The recovered-memory group was more prone to false recognition than was the repressed-memory group, and both of these groups were more prone to false recognition than was the control group. The authors identify false recognition in the DRM paradigm as a type of source-monitoring error. Source monitoring is "remembering how, when, and where a memory is acquired" (Clancy et al., 2002, p. 460). Not remembering the source of information on alien abductions may be related to the creation of false memories of alien abduction (for example, incorrectly remembering an alien abduction as having been a personal experience rather than something watched in a movie years before).

Clancy et al. (2000) compared four groups of women using the DRM paradigm: a group reporting recovered memories of CSA; a group who believed that they had been sexually abused as children but had no memory of the abuse (repressed-memory group); women who had been sexually abused as children and had always remembered it (continuous-memory group); and finally a control group of women with no histories of sexual abuse. Women reporting recovered memories of CSA had a higher rate of false recognition of nonpresented theme words than did women who had never forgotten their abuse; that is, they were more likely to mistakenly report having seen a word that captured the gist of words they had seen. Again, as a type of source-monitoring error, these women may have generated an image of the nonpresented theme word while processing semantically related words and then mistakenly "remembered" having seen the theme word. McNally, Clancy, Barrett, and Parker (2005) refer to this as a *deficit in reality monitoring*—a form of source monitoring in which one must distinguish memories of *perceived* events from memories of *imagined* events. Clancy et al. (2000) recommend using caution when extrapolating their findings to clinical settings but point out that the results are "consistent with the hypothesis that women who report recovered memories of sexual abuse are more prone than others to develop certain types of illusory memories" (p. 30). Are there other characteristics of individuals that make them more susceptible to implanted memories? There is some evidence that more susceptible individuals are more dissociative, introverted, and imaginative than those not as susceptible (e.g., Hyman & Billings, 1998; Porter et al., 1999).

Other comparisons have been made between reports of memories known to be true and those known to be false. Researchers have found statistical differences in some studies but not in others (Loftus, 2004), and much more work remains to be done before it is possible to definitively determine the difference between the two. Studies that have demonstrated distinctions between true and false memories have shown the following: True memories are held with higher confidence than are false memories (Loftus & Pickrell, 1995; Smith et al., 2003); people use more words to describe the true memories (Loftus & Pickrell, 1995); false memories are less coherent than real memories (Porter et al., 1999); and people report false memories as less clear and vivid than are true ones (Loftus & Pickrell, 1995). In one study, there was a trend for the two types of memories to differ in perspective: Participants were more likely to describe implanted memories from a "participant" perspective and to describe true reports from an "observer" perspective (Porter et al., 1999).

SUMMARY

Repressed memories are those that, in Freudian terms, are blocked from consciousness because of their threatening or anxiety-provoking nature. Freud theorized and wrote about repressed memory in the early part of the 20th century. In the last two decades of the 20th century, media sources such as self-help books introduced the notion of repressed memory to the general public. Subsequently, the topic was sensationalized by the media, and many people became convinced that this phenomenon was common (APA, 1995). By the late 1990s and early 2000s, prestigious organizations (such as the American Psychiatric Association) were warning about the problems of false memories being created in therapy, and individuals who had been led to recover false memories in therapy were successfully suing their therapists; public attitude began to change toward recovered memory therapy (Loftus, 2003a).

Before this attitude shift began, however, the notion of repressed and recovered memories in cases of childhood sexual abuse led to one of the most heated controversies in the history of psychology. The debate was not over the existence of CSA; all concerned agreed that CSA occurs all too frequently. (The percentages of adult survivors of CSA in the United States have been estimated at 20 percent of women and 5 to 10 percent of men; Finkelhor & Dzuiba-Leatherman, 1994). The debate was over whether or not traumatic memories could, in fact, be repressed and then recovered years later with no recall in between—that is, could traumatic memories differ so dramatically from nontraumatic memories that they did not conform to the typical forgetting curve? There also was the concern that memories of CSA could be created through suggestion (either in therapy or in the form of self-help books and checklists), and that if false memories were created and then recanted, real survivors of CSA might not be believed (Madill & Holch, 2004). Efforts at helping individuals recover ostensibly repressed memories of CSA led, in some cases, to false-memory reports and subsequent accusations (and sometimes criminal proceedings) against alleged abusers. Controversy raged not only over potentially unethical therapeutic practices but also over whether "[u]ncritical acceptance of every trauma memory report [could] harm the false victims and, also sadly, trivialize the experiences of the true victims [of CSA]" (Loftus, 2003b, p. 871).

Although it is impossible to conduct experimental research on memories of CSA due to ethical constraints, psychological methods have contributed greatly to our understanding of how people remember, how they forget, and how they may come to believe things that never happened. There is no empirical evidence from survey research or experiments on non-CSA memories that traumatic memories differ in any substantial way from nontraumatic memories or that memories can be repressed (Porter et al., 2003). There are case reports indicating that it is possible that people can sometimes recall things in adulthood that had not been thought of for years (Lynn et al., 2003), and these memories are not necessarily false. However, suggestion can lead to false memories and these memories can be expressed with a great deal of confidence, detail, and emotion (Loftus, 2003b)—that is, rich false memories can seem and feel as real as true memories (Loftus & Bernstein, 2005).

Human memory is a reconstructive process; it is malleable and sometimes misleading (Loftus & Bernstein, 2005). Psychologists who have conducted applied memory research have an obligation to dispel common myths of memory and to help reduce social problems associated with the misleading aspects of human memory. Loftus and Bernstein advocate education of three groups of individuals in order to accomplish these goals: education of the general public so that people are less susceptible to the creation of false memories, education of police and other investigators about the power they have to influence memory, and education of clinicians that the use of techniques such as guided imagery to assist clients in getting in touch with so-called repressed memories may be helping them to create, rather than remember, a past (Loftus & Bernstein, 2005).

REFERENCES AND FURTHER READINGS

Alison, L., Kebbell, M., & Lewis, P. (2006). Considerations for experts in assessing the credibility of recovered memories of child sexual abuse: The importance of maintaining a case-specific focus. *Psychology, Public Policy, and Law, 12*(4), 419–441.

Alpert, J. L. (1996). Professional practice, psychological science, and the recovered memory debate. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate* (pp. 325–340). San Diego, CA: Academic Press.

American Psychiatric Association. (2000). *Therapies focused on memories of childhood physical and sexual abuse: Position statement.* Retrieved April 29, 2007, from http://www.psych.org/edu/other_res/lib_archives/archives/200002.pdf

American Psychological Association. (1995). *Questions and answers about memories of childhood abuse.* Retrieved

January 15, 2007, from http://www.apa.org/topics/memories.html

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Bass, E., & Davis, L. (1988). *The courage to heal: A guide for women survivors of child sexual abuse.* New York: Harper & Row.

Berliner, L., Hyman, I., Thomas, A., & Fitzgerald, M. (2003). Children's memory for trauma and positive experiences. *Journal of Traumatic Stress, 16*(3), 229–236.

Braun, K. A., Ellis, R., & Loftus, E. F. (2002). Make my memory: How advertising can change our memories of the past. *Psychology and Marketing, 19,* 1–23.

Clancy, S. A., McNally, R. J., Schacter, D. L., Lenzenweger, M. F., & Pitman, R. K. (2002). Memory distortion in people reporting abduction by aliens. *Journal of Abnormal Psychology, 111*(3), 455–461.

Clancy, S. A., Schacter, D. L., McNally, R. J., & Pitman, R. K. (2000). False recognition in women reporting recovered memories of sexual abuse. *Psychological Science, 11*, 26–31.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* New York: Teacher's College, Columbia University. (Original work published 1885)

Finkelhor, D., & Dzuiba-Leatherman, J. (1994). Victimization of children. *American Psychologist, 49,* 173–183.

Freud, S. (1910). The origin and development of psychoanalysis. *American Journal of Psychology, 21,* 181–218.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin and Review, 3,* 208–214.

Goodman, G. S., Ghetti, S., Quas, J. A., Edelstein, R. S., Alexander, K. W., Redlich, A. D., et al. (2003). A prospective study of memory for child sexual abuse: New findings relevant to the repressed-memory controversy. *Psychological Science, 14*(2), 113–118.

Gore-Felton, C., Koopman, C., Thoresen, C., Arnow, B., Bridges, E., & Spiegel, D. (2000). Psychologists' beliefs and clinical characteristics: Judging the veracity of childhood sexual abuse memories. *Professional Psychology: Research and Practice, 31,* 372–377.

Grinley, M. J. (2002). *Effects of advertising on semantic and episodic memory.* Unpublished master's thesis, University of Washington, Seattle.

Heaps, C. M., & Nash, M. (1999). Individual differences in imagination inflation. *Psychonomic Bulletin & Review, 6,* 313–318.

Heaps, C. M., & Nash, M. (2001). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 920–930.

Hyman, I. E., Jr., & Billings, F. J. (1998). Individual differences and the creation of false childhood memories. *Memory, 6,* 1–20.

Hyman, I. E., Jr., Husband, T. H., & Billings, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology, 9*, 181–197.

Kanter, J. W., Kohlenberg, R. J., & Loftus, E. F. (2002). Demand characteristics, treatment rationales, and cognitive therapy for depression. *Prevention & Treatment, 5*. Retrieved April 24, 2007, from http://www.uwm.edu/~jkanter/demand.pdf

Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science, 15*(3), 149–154.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

Loftus, E. F. (1997). Creating false memories. *Scientific American, 277*(3), 70–75.

Loftus, E. F. (2003a). The dangers of memory. In R. J. Sternberg (Ed.), *Psychologists defying the crowd: Stories of those who battled the establishment and won* (pp. 105–117). Washington, DC: American Psychological Association.

Loftus, E. F. (2003b). Make-believe memories. *American Psychologist, 58,* 867–873.

Loftus, E. F. (2004). Memories of things unseen. *Current Directions in Psychological Science, 13*(4), 145–147.

Loftus, E. F. (2005a). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12,* 361–368.

Loftus, E. F. (2005b). Searching for the neurobiology of the misinformation effect. *Learning & Memory, 12,* 1–2.

Loftus, E. F., & Bernstein, D. M. (2005). Rich false memories: The royal road to success. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 101–113). Washington, DC: American Psychological Association.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118,* 100–104.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning, 4,* 19–31.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25*, 720–725.

Lynn, S. J., Lock, T., Loftus, E. F., Krackow, E., & Lilienfeld, S. O. (2003). The remembrance of things past: Problematic memory recovery techniques in psychotherapy. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 205–239). New York: Guilford Press.

Madill, A., & Holch, P. (2004). A range of memory possibilities: The challenge of the false memory debate for clinicians and researchers. *Clinical Psychology and Psychotherapy, 11,* 299–310.

Mazzoni, G. A. L., Loftus, E. F., & Kirsch, I. (2001). Changing beliefs about implausible autobiographical events: A little plausibility goes a long way. *Journal of Experimental Psychology: Applied, 7,* 51–59.

McNally, R. J. (2003). Recovering memories of trauma: A view from the laboratory. *Current Directions in Psychological Science, 12*(1), 32–35.

McNally, R. J. (2004). Is traumatic amnesia nothing but psychiatric folklore? *Cognitive Behaviour Therapy, 33*(2), 97–101.

McNally, R. J., Clancy, S. A., Barrett, H. M., & Parker, H. A. (2005). Reality monitoring in adults reporting repressed, recovered, or continuous memories of childhood sexual abuse. *Journal of Abnormal Psychology, 114*(1), 147–152.

Ost, J. (2003). Essay book review: Seeking the middle ground in the 'memory wars.' *British Journal of Psychology, 94*(1), 125–139.

Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science, 8,* 437–441.

Pezdek, K., & Hodge, D. (1999). Planting false childhood memories in children: The role of event plausibility. *Child Development, 70,* 887–895.

Polusny, M. A., & Follette, V. M. (1996). Remembering childhood sexual abuse: A national survey of psychologists' clinical practices, beliefs, and personal experiences. *Professional Psychology: Research and Practice, 27,* 41–52.

Poole, D. A., Lindsay, D. S., Memon, A., & Bull, R. (1995). Psychotherapy and the recovery of memories of childhood sexual abuse: U. S. and British practitioners' opinions, practices, and experiences. *Journal of Clinical and Consulting Psychology, 63,* 426–437.

Porter, S., Campbell, M. A., Birt, A. R., & Woodworth, M. T. (2003). "He said, she said": A psychological perspective on historical memory evidence in the courtroom. *Canadian Psychology, 44*(3), 190–206.

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior, 23*, 517–537.

Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past.* New York: Basic Books.

Smith, P. T. (2000). A jigsaw puzzle theory of memory. *Memory, 8*(4), 245–264.

Smith, S. M., Gleaves, D. H., Pierce, B. H., Williams, T. L., Gilliland, T. R., & Gerkens, D. R. (2003). Eliciting and comparing false and recovered memories: An experimental approach. *Applied Cognitive Psychology, 17*, 251–279.

Taylor, S. (2004). Amnesia, folklore, and folks: Recovered memories in clinical practice. *Cognitive Behaviour Therapy, 33*(2), 105–108.

Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory and Cognition, 30*, 423–431.

Wade, K. A., Garry, M., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin & Review, 9*, 597–603.

Wilsnack, S. C., Wonderlich, S. A., Kristjanson, A. F., Vogeltanz-Holm, N. D., & Wilsnack, R. W. (2002). Self-reports of forgetting and remembering childhood sexual abuse in a nationally representative sample of US women. *Child Abuse & Neglect, 26*, 139–147.

49

Language and Language Development

David Kreiner
University of Central Missouri

It seems easy for most people to say something that comes to mind, to understand the words coming out of another person's mouth, or to read the words in a sentence like this one. We hardly even think about what our brains have to do in order to understand and produce language. However, our language abilities are so sophisticated that even the most powerful computers cannot match the language abilities of a five-year-old child. What makes it so easy for children to learn language? What *is* language?

The study of how we understand and produce language is an important part of modern psychology. In the years after World War II, there was a renewed interest in studying the mind. The previous decades, especially in American psychology, were dominated by the behaviorist perspective, which excluded the study of the mind per se. One spark that helped lay the groundwork for a new perspective was the argument that language abilities cannot be explained without studying the mind (Chomsky, 1959). This argument helped convince many psychologists that it was both acceptable and worthwhile to study mental processes in addition to overt behavior.

The questions of what language is and how it works have captured the imaginations of scholars in diverse fields including psychology, linguistics, philosophy, computer science, and anthropology. Many people feel that language is at the core of what it means to be human. This chapter begins by discussing how language can be defined. We will then review some of the methods used to study language abilities. Three central issues in the psychology of language will then be summarized: how language evolved, how language is acquired, and the relationship between language and thought. We will conclude by pointing out how the psychology of language can be applied to our lives.

WHAT IS LANGUAGE?

Language is a way of communicating, but it is more than that. There are many ways of communicating other than using language. We can communicate with one another through art, music, and nonverbal behaviors such as facial expressions. What makes language different from these other communication systems? George Miller (1990) pointed out that language is the only communication system that is also a representational system. In other words, language allows us both to represent the world in our minds and to talk about those representations of the world with one another.

Properties of Language

Three defining properties of language are that it is hierarchical, rule-governed, and generative. *Hierarchical* means that language is composed of small units that can be combined into larger units, which can themselves be combined into larger units, and so on. Speech sounds can be combined into words, words can be combined into phrases, and phrases can be combined into sentences.

The way that the units are combined is not arbitrary, but is regular and *rule-governed*. The speakers of a language follow these rules even if they cannot explain exactly what the rules are. Any competent speaker of English knows that "The big cat ate the pancake" is a sentence, and that "Cat big the the pancake ate" is not a sentence. But most of us cannot state exactly what rules were followed in the first example and broken in the second example. We know the rules of language tacitly and follow them without having to think about them.

Language is *generative* in the sense that a relatively small number of units can be combined into a much larger number of possible messages. In fact, it is easy to show that there is an infinite number of possible sentences in any language. A new sentence can be generated from an existing sentence by adding another phrase, as in the following:

"The cat ate the pancake."
"The cat ate the pancake and the waffle."
"The cat ate the pancake and the waffle and the egg."

We could continue adding phrases indefinitely, which demonstrates that there is an infinite number of possible sentences that can be produced by any speaker of the language. The generative property of language allows us to produce new sentences, including ones that have never been heard before, and still be understood. Therefore, it is impossible to learn a language by simply memorizing all the sentences. A language user must know the units that make up the language (such as speech sounds and words) as well as the rules for combining the units.

To determine whether a system for communicating is truly a language, we need to determine whether the system is hierarchical, rule-governed, and generative. Consider the use of gestures such as moving one's hands to communicate. We often use gestures while we are talking to convey additional information to listeners. However, gestures do not have the properties of language when they are used in combination with spoken language. Gestures that are used in combination with speech are not made up of units that can be combined into larger units. Interestingly, though, when gestures are used on their own, without speech, to communicate, as in sign language used by hearing-impaired individuals, the gestures do have the properties of language (Goldin-Meadow, 2006).

Language Processes

Language is not a single process. This chapter will focus mostly on spoken language, but it is important to keep in mind that language can make use of other modalities, too. For example, language can be communicated in gestural form (such as American Sign Language). As you read this chapter, you are processing language in its written form. There are important differences among these modalities (speech, sign language, and written language).

For each of these modalities, scientists who study language investigate both the production and comprehension of language. When studying spoken language, we are interested in both speech production and in how listeners perceive and understand speech. When studying written language, we are interested in how writers translate their ideas to a series of ink marks on a page as well as how readers interpret those ink marks.

The Structure of Language

Our ability to produce and perceive language is remarkable when we consider the problems that confront our brains. The hierarchical structure of language means that language users must be able to understand and use several types of information just to understand a single sentence. In order to comprehend the spoken sentence "The cat ate the pancake," the listener must be able to process information about the speech sounds, the meanings of the words, the grammatical structure of the sentence, and the goal of the speaker in uttering the sentence.

The speech sounds that make up words are called *phonemes*. The word "cat" comprises three phonemes. (Phonemes should not be confused with the letters used to spell the word, as we are considering a spoken sentence in this example.) The three phonemes roughly correspond to the sounds *kuh*, *ah*, and *tuh*. Each phoneme is actually a category of sounds because we perceive a variety of acoustically different sounds as the same phoneme. The *kuh* in "cat" is physically different from the *kuh* in "pancake," even though most English speakers perceive them as the same. The acoustic information in a phoneme can vary dramatically depending on the word it is in, the voice of the speaker, and the rate of speech. Despite these differences, listeners perceive all the examples as one speech sound. The ability to accomplish this feat of classification is one reason why computer speech recognition software is inferior to humans' ability for speech recognition.

Correctly perceiving phonemes is important for understanding speech, but it is not enough. The listener must be able to combine the phonemes into words and understand what those words mean. The meanings of words, phrases, and sentences represent the *semantic* level of language. Many words are made up of multiple parts, or *morphemes*, which carry meaning. For example, the word "cooked" is made up of two morphemes: "cook" and "ed," with the suffix "ed" indicating that the cooking took place in the past.

To understand what someone is saying, we need to do more than just understand the individual morphemes and words. We need to understand the *syntax*, or grammatical structure, of the phrases and sentences. Consider the following sentence:

"His face was flushed but his broad shoulders saved him."

The word "flushed" can be understood two different ways, depending on how the listener processes the syntax. If the listener understands "flushed" to be an adjective that describes the color of the person's face, then the rest of the sentence is difficult to understand. The listener is left trying to understand why the person's broad shoulders might have saved him from being embarrassed. But the word "flushed" can also be understood as the past tense of the verb "flush." Interpreting the syntax in this way completely changes the meaning of the sentence, as it is now obvious that the person's shoulders saved him from going down the toilet. Note that the phonemes, morphemes, and words are identical, but the meaning is different depending on the listener's understanding of the syntax.

Even with an understanding of the syntax, the listener or reader can sometimes misunderstand what the speaker or writer intended. Language users must make use of *pragmatics*, or understanding the purpose of the communication. The question "Can you sit down?" may be interpreted as a request for information about the abilities of the listener. If a doctor asks this question of a patient visiting her office with concerns about joint pain, the patient might conclude that the purpose of the question is to determine whether the patient is capable of sitting. An appropriate response might then be, "Yes, but my knees hurt when I sit for a long time." Suppose instead that the question "Can you sit down?" is asked by a spectator at a baseball game, directed at an individual standing in front of her. In this context, the listener might interpret the question to mean, "I can't see the ball game because you're in my way. Please sit down so I can see." The listener's response might be to sit down. Note that in both examples the phonemes, morphemes, words, and syntax are the same, but the meaning of the utterance still differs.

METHODS FOR STUDYING LANGUAGE PROCESSING

The study of language is multidisciplinary, with scientists from many different fields investigating language using different methods. The diversity of techniques that have been used to investigate language has resulted in a rich research literature on language processing.

Experimental Studies

Psychologists often use different types of experimental designs to determine what factors influence particular behaviors. This approach has been helpful in testing theories and models of language processing, which often yield predictions about factors that should influence how quickly or accurately individuals respond in language tasks. For example, some theories of reading predict that high-frequency words (those that are very common, such as "dollar") should be recognized more quickly than low-frequency words (those that are rarely encountered, such as "vestal"). Numerous experiments have supported this prediction (e.g, Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Such evidence has been helpful in evaluating theories that explain how we recognize words.

Computational Modeling

A more recent tool for testing theories of language is computational modeling, or the use of computer simulations. The idea is to implement the major claims of a theory in computer software, and then test the software to see if it produces results similar to the results found in experimental studies with human participants. Coltheart et al. (2001) demonstrated that their theory of reading, when implemented as a computer model, produced results similar to a wide range of experimental studies with humans. For example, the computer model has difficulty with the same types of words that human readers find more difficult to recognize. The similarity in the pattern of performance between the computer model and the human data shows that the model is a plausible one. A particular advantage of computer modeling is that it forces the theorists to implement their ideas in very specific ways and allows for clear tests of what the model claims. In scientific psychology, a model that yields clear, testable predictions is always preferable to a model that is too vague to be tested decisively.

Studies of Brain Activity

Advances in technology in recent decades have allowed scientists to study patterns of brain activity related to language processing. It is now possible to learn how the brain processes language by measuring brain activity in individuals as they engage in various language tasks. Brain-imaging technology such as the PET scan has demonstrated, for example, that different brain areas are active when individuals read words and when they listen to words. Researchers have also used electrodes to stimulate portions of the cortex as a method of determining which brain areas are important for language; this technique is called Cortical Stimulation Mapping (J. G. Ojemann, G. Ojemann, Lettich, & Berger, 1989). A variety of brain-imaging technologies can be used to compare individual differences in brain structure to performance in language tasks.

Studies of Language Deficits

One way to find out how a complicated ability such as language works is to study what goes wrong when individuals suffer from impaired abilities. The strategy is to determine what particular abilities have been impaired by damage to the brain and to match those impairments to specific brain areas. When this matching can be done, it provides good evidence that a particular brain area is important for a particular language ability.

Two well-known examples are Broca's aphasia and Wernicke's aphasia (Caplan, 1994). *Broca's aphasia* is characterized by difficulty producing speech and is associated with damage to an area in the frontal lobe that is now called Broca's area. An individual suffering from Broca's aphasia typically has no difficulty understanding language. The individual knows what he or she wants to say, but has great difficulty in producing fluent speech. Broca's area appears to play an important role in the production of speech.

Individuals who have *Wernicke's aphasia* can speak fluently but have difficulty understanding speech. As a result, these individuals may produce sentences that are grammatical but that do not seem to mean anything. Wernicke's aphasia results from damage to a particular area in the temporal lobe that is now called Wernicke's area. We can infer that Wernicke's area plays an important role in helping us comprehend speech.

The location of these areas was first proposed by physicians who performed autopsies on individuals who had demonstrated these impairments. More recently, the locations of Broca's area and Wernicke's area have been documented by recording brain activity in living patients, with techniques such as fMRI (functional magnetic resonance imaging).

Analysis of Language Samples

Some of the simplest but most informative evidence about language comes from the analysis of patterns in language samples. Patterns of errors in speech can be revealing about the underlying mental processes that must have produced them. Saying "the nipper is zarrow" instead of "the zipper is narrow" indicates that individual phonemes can be switched between words (Fromkin, 1973). Professor Spooner of Oxford University is reputed to have made a habit of switching phonemes, resulting in amusing sentences such as "You have tasted the whole worm" instead of "You have wasted the whole term." The linguistic structure of the sentence must be planned before the first word comes out of the speaker's mouth. If this planning does not take place, it is difficult to explain where the "n" sound came from in "the nipper is zarrow." It seems obvious that it was borrowed from the beginning of "narrow," a word that occurs later in the sentence.

A second example of the use of language samples to provide evidence about language processing is the analysis of writing samples. Pennebaker and his colleagues (e.g., Pennebaker & Graybeal, 2001) reported that patterns of word use in language samples can predict differences in health, academic performance, and a variety of behaviors. For example, Pennebaker and colleagues found that greater use of cognitive words (e.g., "know" or "understand") in writing samples is a good predictor of health improvement.

Biological and Anthropological Research

Language is a biological phenomenon, and valuable lessons can be learned when it is approached from this perspective. Biological and anthropological evidence can be particularly informative about how language abilities evolved. In many respects, we are similar to our primate cousins, but there are noticeable differences in language use. Although other primates sometimes communicate through vocalizations, they do not talk, at least not in anything resembling human language.

Why is it that primates such as chimpanzees cannot be taught to speak? Part of the answer is that their vocal tracts do not allow them to produce the wide range of speech sounds that are included in human languages. A key biological difference between humans and other primates is the location of the larynx (or "voicebox") in the throat. In humans, the larynx is located lower in the throat, which allows for a much wider range of sounds to be produced. The cost of this anatomical difference is that humans are at much higher risk of choking to death on their food. The evolutionary advantages of human language must be powerful enough to overcome the disadvantage that arises from the risk of choking. Lieberman (2000) pointed out that our ancestors who benefited from this anatomical change in the vocal tract must have already had language abilities. This species could not have benefited from the lowered larynx unless they already had language abilities that could be improved with the altered vocal tract.

THE EVOLUTION OF LANGUAGE

An important and controversial question about language is how language abilities could have evolved. The ability to use the same system both to represent the world and to communicate about past, present, and future situations clearly gave our species enormous advantages in surviving and reproducing.

We can get some clues about how language evolved by studying living species that are similar to us, such as chimpanzees. A common misconception is that modern humans evolved from modern chimpanzees or monkeys. This progression is clearly not the case; instead, we share common ancestors with other living species. We can learn what physical and behavioral characteristics our common ancestors had by comparing humans to other species. If two genetically similar living species share a particular characteristic, it is likely that their common ancestor also shared that characteristic. A species does not simply develop a behavior such as language out of thin air; instead, it must evolve from precursor abilities that were present in ancestral species.

One of the differences between humans and other primates is that humans have much greater control over the face and mouth, allowing us to produce a wide variety of sounds and to combine these sounds into rapid sequences. These differences can be traced to a specific gene called FOXP2 (Marcus & Fisher, 2003). Although many other animals have the FOXP2 gene, the human version differs from that in other animals, and this difference is only about 200,000 years old. It is certainly not the case that FOXP2 is *the* language gene in the sense that it is solely responsible for our language capability. It is, however, a useful clue about how language abilities may have evolved. The change in this particular gene may have allowed our ancestors to develop their existing vocal repertoire into considerably more complex and useful forms of communication.

Other living primate species also communicate with vocal signals. For example, vervet monkeys use alarm calls that alert other vervet monkeys to the presence of a predator. However, vervet monkeys use these alert calls equally for their companion monkeys who are already aware of the presence of the predator and those who are not aware of the predator. In contrast, humans take into account the mental states of the individuals with whom they are speaking. However, some primates show an apparent precursor to this pragmatic use of language; they change their vocalizations depending on their own social status (Zuberbuhler, 2005).

In terms of speech perception, there are also interesting similarities to and differences from other species. Recall that a phoneme is really a category of speech sounds: The *kuh* sound in "cat" is physically different from the *kuh* sound in "pancake," but humans readily perceive the two sounds as the same phoneme. This ability is referred to as *categorical perception*. Other species (including chinchillas and quail) can learn to perceive phonemes in specific categories. An interesting difference is that other species require large amounts of systematic training and feedback in order to show categorical perception of phonemes, while human infants require little or no training. This difference suggests that, although other species may be able to acquire some language abilities, the mechanisms for acquiring them are not the same.

We have focused mainly on the evolution of spoken language. It is worth noting that written language abilities are very different in terms of evolutionary history. Writing systems appear to date back only about 6,000 years. For most of history, then, very few individuals have been able to read and write. These facts indicate that the human brain could not have evolved specific abilities for written language. Instead, we must rely on other cognitive abilities, which may explain why humans are so much more successful at acquiring spoken language than at learning to read and write.

IS LANGUAGE LEARNED OR ACQUIRED?

A central issue in the psychology of language is the question of how we acquire our language abilities. The mystery is that children become proficient in producing and understanding speech very early in life and without formal education. Children do not have to be taught how to speak; they show up for the first day of kindergarten with language abilities approaching those of adults.

Early language development proceeds in a very similar way for different children. Development of speech ability follows a regular sequence, from babbling beginning at about six months of age, to single words by about one year, and then longer utterances. The precise age of each development does vary across children, but the sequence is similar. This developmental sequence is also similar across different languages.

Speech perception abilities also develop in a regular sequence. Infants are born with general speech perception abilities that apply to all languages; they can perceive distinctions between phonemes that occur in any language. However, by about age one, infants have tuned their speech perception abilities to the language spoken in their environment, losing the ability to perceive distinctions in other languages but retaining the ability to distinguish phonemes in their own language (Juscyzk, 1997).

The speed with which most children acquire language is remarkable. Children learn an average of about eight words per day between the ages of 18 months and 6 years. How do children learn words so quickly? The answer is that there is more than one way to learn words. Children make use of several types of information. Younger infants rely mainly on perceptual salience: They associate a word with the most interesting object that they can see. By age two, children rely much more on social cues such as associating a word with an object at which a caregiver's gaze is directed (Golinkoff & Hirsh-Pasek, 2006).

One of the key problems we face in learning words is being able to segment speech into individual words. As anyone can attest who has listened to people speaking an unfamiliar foreign language, it is very difficult to determine where one word ends and another begins. How do children learn where the word boundaries are? Children appear to take advantage of statistical patterns in speech. For example, infants can take advantage of the fact that the syllable *ty* is reasonably likely to follow the syllable *pre*, but *ba* is very unlikely to follow *ty* (Saffran, 2003). Thus, when hearing a sentence such as "What a pretty baby," the infant is more likely to perceive a word boundary between "pretty" and "baby" than in the middle of "pretty."

The ability to produce and understand grammatically correct sentences develops as vocabulary is acquired. As children acquire language, they do not do so by simply memorizing lots of sentences. Children often produce sentences that their parents and caregivers never utter, and they can understand sentences that they have never heard before. The generative, hierarchical, and rule-governed properties of language come naturally to children.

We know that children learn a language by learning its rules because they sometimes overgeneralize those rules. In English, the general rule for making a verb past tense is to add the morpheme *–ed* to the end of the verb. The past tense of "walk" is "walked" and the past tense of "ask" is "asked." But some verbs in English are exceptions; they do not follow the rules. The past tense of "run" is "ran" and the past tense of "eat" is "ate." As children begin to use these irregular verbs, they typically use the verb correctly: "I ran away." The same child who earlier used "ran" correctly will later learn the past-tense rule but will begin to overgeneralize it, producing sentences such as "I runned away." Eventually, with experience, the child will learn which words are exceptions from the rule and will go back to using "ran."

It is more accurate to say that children acquire language than that they learn it. Children spontaneously acquire language as long as they are neurologically normal and in an environment with language. They acquire language rapidly and show evidence of learning the underlying structure of the language, not just memorizing sentences. They follow a similar developmental sequence, which suggests that we come into the world biologically prepared to acquire language quickly.

These facts do not imply that we are born with specific knowledge about our language or even about human languages as a group. Saffran (2003) suggested that common patterns of language development may instead be due to constraints on learning. As noted previously, children can make use of statistical properties in speech as they acquire the language. Saffran argued that some statistical properties are easier to learn than others, and those statistical properties that are easiest for our brains to learn are more likely to be included in the world's languages. Languages that included patterns that were more difficult to learn could not have become popular. In other words, it may be that human languages have evolved so that they are easy for children to learn.

LANGUAGE AND THOUGHT

Our language abilities clearly depend on a variety of cognitive processes such as attention, pattern recognition, memory, and reasoning. In order to understand a spoken sentence, the listener must pay attention to what the speaker is saying, recognize the phonemes and words, relate those words to pre-existing knowledge, and make reasonable inferences about what the speaker is trying to communicate.

Language abilities rely on these cognitive processes, but to what extent are these cognitive processes influenced by language? This question is a central issue for the psychology of language as well as for cognitive psychology.

Benjamin Whorf was an inspector for an insurance company as well as an amateur linguist. Whorf wondered whether some accidents that he investigated might have been a result of language limiting the way that individuals could think about a situation. In one case, a fire had occurred in a warehouse after an employee had discarded a match into a gasoline drum. A sign near the drum had indicated EMPTY. Whorf thought that the employee's understanding of the word "empty" might have limited the way the employee could think about the situation. If the drum is empty, then nothing is in it, and it is safe to discard a match. The reality, of course, was that the "empty" drum still contained gasoline vapors, which ignited on contact with the match. Based on observations such as this one, in addition to his study of various languages, Whorf proposed that language can limit or determine thought (Carroll, 1956). This idea, referred to as the *linguistic relativity hypothesis*, has been the subject of considerable research.

One way to test the linguistic relativity hypothesis is to find a way in which languages differ from one another and then find out whether the speakers of those languages differ in how they think. For example, languages differ in the number of words for different colors. In the Dani language, spoken by natives of New Guinea, there are only two color words: *mola* for bright colors and *mili* for dark colors. Dani speakers do not have separate words for blue, green, red, yellow, and so on. If Whorf was correct that language determines thought, then we should expect Dani speakers to differ in how they categorize colors when compared to English speakers. The evidence on this point is mixed. Early studies indicated no differences. Rosch (1973) reported that both Dani and English speakers have better memory for focal colors or the "best" examples of a color category. Most people will agree that a similar shade of red is the "best" red. That shade is a focal color, and Rosch's study suggested that it was easier to recall having seen that shade than other, nonfocal shades. The crucial point is that this finding was also true for Dani speakers, whose language does not have a specific word for the color red.

More recent studies have reported differences in color categorization depending on language. Roberson, Davies, and Davidoff (2000) studied speakers of the New Guinea language Berinmo, which has five color words that differ from the color words in English. Both English speakers and Berinmo speakers are better at perceiving differences between colors that cross color boundaries in their own language. For example, blue and green are represented by the same basic color word in Berinmo. English speakers are better at discriminating shades of blue from shades of green than shades within either blue or green, while Berinmo speakers did not show this advantage. However, Berinmo speakers are better at telling the difference between colors that are represented by different words in their language. It is not the case that Berinmo speakers or English speakers are better at discriminating between colors. Instead, this research shows that the colors that are easiest to discriminate depend on what language the person speaks. This type of finding is consistent with the linguistic relativity hypothesis.

The linguistic relativity hypothesis can be characterized as having a strong form and a weak form. The weak form of the hypothesis is that language can influence the way that we think. In some ways, this point is obvious. After all, there is rarely a reason to use language except to influence the way that other individuals think.

The strong form of the hypothesis is that language limits or determines what we think. Are speakers of Dani even able to think about the difference between colors such as blue and green, given that they do not have separate color words for them? Rosch's research suggests that they can think about these colors even without having words for them. We often think about concepts for which we do not have particular words. Almost everyone is familiar with the awkward situation in which two people are walking down a hallway from opposite directions, notice that they are on a collision path, and take a step to one side. Both individuals happen to step the same way, leaving them on a collision path again. This dance sequence may be repeated several times, and finally the two individuals can pass each other, perhaps both smiling. We can understand this situation without having a particular word to describe it. Thus, the strongest form of Whorf's hypothesis is untenable. We can, however, reasonably conclude that language affects cognitive processes.

APPLICATIONS OF THE PSYCHOLOGY OF LANGUAGE

Nearly everything we do in our daily lives has something to do with how we use language. An understanding of the psychology of language can be useful in helping us think about many aspects of modern life.

Teaching of Language

What we know about acquisition of language skills can inform our decisions about teaching language to children. Children's abilities to speak and to understand speech develop naturally, without the need for formal instruction. Learning spoken language is more like learning how to walk than learning how to ride a bicycle. Children do not really learn how to walk (or how to speak); instead, they acquire these abilities as they develop biologically. The regular sequence of development across children and across cultures and the development of these abilities in children who do not receive any formal instruction demonstrate that it is not necessary to teach children how to walk or how to speak. Biologically normal children, in reasonably normal environments, just do it. Of course, this naturally occurring developmental process does not prevent many parents and caregivers from attempting to teach spoken language to children. Parents may believe that, because they have made special efforts to teach their children how to speak, and because the children did learn how to speak, their instructional efforts were necessary for proper language development. This relationship is illusory, as children whose parents do not make such efforts still acquire spoken language abilities. Children are notoriously resistant to correction when it comes to language. Recall that children typically learn language rules and overgeneralize them such as saying "runned" instead of "ran." Caregivers often find it excessively difficult to stop children from making these sorts of mistakes.

It is useful for parents and caregivers to understand that children are typically capable of understanding more than they can say. In the normal developmental sequence, children are able to understand sentences before they are able to produce sentences of the same grammatical sophistication. A sentence such as "The pancake that I saw in the kitchen was eaten by the cat" may be easily understood by a two-year-old, but that same child is unlikely to produce sentences with that level of grammatical complexity.

Another interesting application of the research literature on language acquisition is what it can tell us about learning a second language. We are born with the capacity to acquire any human language. The same abilities that make it easy for us to acquire a native language can make it more difficult to learn a second language later in life. Consider the development of our speech perception abilities. Newborn infants are capable of discriminating among phonemes in any human language. By age one, we lose the ability to perceive distinctions between phonemes that are not in our language, which makes it challenging to learn a second language that has a different set of phonemes. In English, there is a distinction between the sounds *luh* and *ruh*. In Japanese, there is no such distinction, making it difficult for a native speaker of Japanese to tell the difference between words such as "rip" and "lip," which differ in only the initial phoneme. It is possible for an adult to learn such distinctions, but it appears to require substantial amounts of practice, which presents a challenge for individuals who wish to learn to speak a second language without an "accent."

Learning to Read

Although we have not addressed in any depth the processes underlying reading skill, we can make several points about learning to read. First, learning to read is a qualitatively different process from acquiring spoken language. Written language is so recent in our history as a species that we could not have evolved special abilities that would help us learn to read. In some ways, learning to read is unnatural. There is no particular reason why a pattern of ink marks on a page should correspond to particular words and sentences. Many children—and adults—fail to learn to read at a fluent level. Those who do learn to read fluently almost always do so as a result of a considerable amount of formal instruction and practice.

Languages differ substantially in how they use written characters to represent spoken language. English is written using a system that is primarily *alphabetic*, meaning that written symbols represent speech sounds. Other languages, such as Chinese, rely on *ideographic* writing systems in which written symbols can represent words or concepts. Still other languages such as one of the Japanese writing systems are *syllabic*: written symbols represent syllables. Depending on the language, the beginning reader must not only figure out the type of writing system involved but also master the correspondences between written symbols and phonemes, syllables, or words.

Simply being able to tell the difference between different letters in an alphabet presents a problem for which our brains are not well prepared. The letters *d* and *b* are mirror images of each other. Physically, they are the same, just in different orientations. For almost everything else in our environments, an object is the same whether it is pointed to the left or to the right. A bird flying to the south does not become a different bird merely by turning around and flying north. But a *d* becomes a different letter, representing a different sound, just by facing a different direction than a *b*.

We should not be surprised that learning to read is difficult and time-consuming compared to the acquisition of speech, which is rapid and almost effortless. This fact has important consequences for how we design our educational systems and for how we interpret difficulties in learning language.

What Constitutes a Language?

We have addressed the question of what makes a language different from any other system of communication. This knowledge can be helpful in addressing misconceptions about what is and is not a "proper" language.

Does sign language count as a language, or is it simply a spoken language that has been translated from sounds into gestures? Many people are familiar with finger-spelling, in which letters are translated to specific configurations of the fingers so that words can be spelled out. Finger-spelling is a slow and inefficient way of communicating, and it does not constitute a separate language. However, finger-spelling is not the primary way that individuals communicate through sign language. True sign languages are actually distinct languages, in the same way that French is a different language from English. One example is American Sign Language (ASL). In ASL, the same levels of language structure exist as in any spoken language, with the exception that signs take the place of phonemes. ASL has syntax, semantics, and pragmatics. The same properties that define spoken language also apply to ASL and other sign language systems: They are generative, hierarchical, and rule-governed. ASL signs are not simply translations of English words. The developmental sequence in which children acquire sign language parallels the developmental sequence for spoken language. Children who are acquiring sign language even "babble" with their hands in the same way that children acquiring spoken language babble by repeating syllables such as *ba ba ba ba.*

One difference between spoken language and sign language is that signs often bear a resemblance to the words or concepts that they represent. In spoken language, the sound of a word has nothing to do with what the word means, with a few exceptions known as onomatopoeia (e.g, the word "meow" sounds something like a cat meowing). There is nothing about the sound of the word "cat" indicating that it represents a cat; speakers of English simply must learn that association. In contrast, the ASL sign for "cat" is a gesture indicating the cat's whiskers.

Despite differences between spoken and signed language, it is misleading to suggest that sign language is not a "real" language. Children who master both a spoken language such as English and a sign language such as ASL are bilingual. Individuals who communicate primarily through sign language should not be considered somehow deficient in their language abilities, or lacking a real language, any more than an English speaker who is not proficient in French should be considered deficient.

Similar points can be made about speakers of dialects within a language. An interesting and controversial example is how many people perceive and react to black English vernacular (BEV), sometimes referred to as Ebonics. Black English is sometimes assumed to be merely a lazy version of English in which certain sounds and grammatical endings are routinely dropped. A similar argument is that black English is a degraded version of standard English and that the use of it portends the decline of the English language in general.

Such beliefs were the basis of a controversy in the Oakland, California, school district in the 1990s. The school district was criticized for a plan to train teachers in how to understand Ebonics so that they could help their students learn standard American English (SAE). However, arguments that BEV is a degraded version of English are not accurate and ignore what we know about the structure and function of language. Pinker (1994) described how BEV is, in fact, just as grammatical, rule-governed, and regular as SAE. One misconception is that speakers of black English leave out more words and word endings in their speech, but this point is not correct. In any language, the rules transform over time, often becoming more efficient rather than "lazier." A good example in SAE is the use of contractions such as "I've" instead of "I have." Although there are some patterns in BEV in which a word may be regularly omitted, there are also patterns in SAE in which a word is regularly omitted that is left intact in BEV. Pinker pointed out that, in BEV, the sentence "He be working" is regular and grammatical, with the word "be" making a meaningful distinction. The inclusion of "be" in this sentence indicates that the person generally does work, that is, has regular employment. In contrast, "He working" in BEV indicates that the person is working right at the moment. This distinction is not made in SAE, where "He's working" could be interpreted in either of these two ways. Does this example indicate that SAE is lazier, or a more degraded form of English, than BEV? No, nor is it reasonable to conclude that BEV is less regular or meaningful than SAE.

Artificial Language Systems

Much of the modern interest in the psychology of language grew out of attempts to create artificial systems for understanding or producing language. For example, researchers trying to build reading machines for the blind found that it was very difficult to get the machines to produce fluent, comprehensible speech. This difficulty led to an explosion of research on speech perception and speech production. The difficulties of constructing such systems have helped us understand just how sophisticated and difficult language processing is, and how remarkable it is that our brains do it so well.

Clearly there have been advances in recent years. Software that takes dictation and translates it into typed words is used widely. Voice-activated customer service systems are no longer a rarity, and they function effectively in most cases. Two aspects of these recent technological advances are illuminating. First, it took many decades and the investment of huge amounts of resources to make such systems function effectively. Second, these systems still do not approach the language abilities of humans. These artificial language systems are successful only because they are tailored to specific situations and types of language. The software that allows you to dictate a paper to your computer is not capable of understanding those sentences and summarizing their meaning. Voice-activated customer service systems are at a loss if the customer says something that is unexpected, but a human customer service agent can understand virtually anything that the customer might say. These two points highlight the amazing language capabilities of the human brain.

SUMMARY

We have learned that language is a fascinating ability that is unlike any other type of communication. Language has been studied from a variety of perspectives with a variety of methods, yielding interesting conclusions about the nature of language, how language might have evolved, how language abilities are acquired, and how language might affect thought.

A key theme is the extent to which language relies on mechanisms in the brain that are specific to language, as opposed to relying on mechanisms that are shared with other cognitive abilities. On one hand, it is apparent that our brains are exceptionally prepared to master a complex communication system that is, to date, beyond the reach of even the most advanced technology. Spoken language (but not written language) is learned with ease and without the need for any formal instruction. On the other hand, many of the problems that the brain confronts in using language are problems that also occur in other aspects of cognition. For example, consider the problem of categorical perception: We perceive physically different sounds as belonging to the same phonemic category. We must solve a similar problem when we learn to categorize a variety of dogs as belonging to same concept of dog, or when we perceive physically different sounds as being examples of the same musical note. The fact that language presents difficult problems for the brain to solve and that the brain in fact solves these problems, does not by itself indicate that language is a special ability, distinct from other human abilities. Nor does the existence of similar problems for other abilities demonstrate that the same underlying cognitive skill is shared; it is possible that the same problem, such as categorical perception, could be solved with different mechanisms.

It is clear that much remains to be learned about our language abilities. Scientists have made substantial progress over the past half-century in understanding how language works. Still, what George Miller noted in 1990 remains true today: "[W]e already know far more about language than we understand" (p. 7).

REFERENCES AND FURTHER READINGS

Caplan, D. (1994). Language and the brain. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1023–1053). San Diego, CA: Academic Press.

Carroll, D. W. (2004). *Psychology of language* (4th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Carroll, J. B. (Ed.). (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Boston: MIT Press.

Chomsky, N. (1959). Review of *Verbal Behavior* by B. F. Skinner. *Language, 35*, 26–58.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Fouts, R. (1997). *Next of kin: What chimpanzees have taught me about who we are*. New York: William Morrow.

Fromkin, V. A. (1973). *Speech errors as linguistic evidence.* The Hague, Netherlands: Mouton & Co.

Goldin-Meadow, S. (2006). Talking and thinking with our hands. *Current Directions in Psychological Science, 15*, 34–39.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science, 15*, 30–33.

Juscyzk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kellogg, R. T. (1999). *The psychology of writing*. New York: Oxford University Press.

Lieberman, P. (2000). Human language and our reptilian brain. *Biology and Medicine, 44*, 32–50.

Marcus, G. F. (1996). Why do children say "breaked?" *Current Directions in Psychological Science, 5*, 81–85.

Marcus, G. F., & Fisher, S. E. (2003). FOXP2 in focus: What can genes tell us about speech and language? *Trends in Cognitive Sciences, 7*, 257–262.

Miller, G. A. (1990). The place of language in a scientific psychology. *Psychological Science, 1*, 7–14.

Ojemann, G., Ojemann, J. G., Lettich, E., & Berger, M. (1989). Cortical language localization in left, dominant hemisphere: An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery, 71*, 316–326.

Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science, 10*, 90–93.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: William Morrow.

Posner, M. I., Petersen, S. E., Fox, P. I., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science, 240*, 1627–1631.

Rayner, K., & Pollatsek, A. (1994). *The psychology of reading*. Mahwah, NJ: Lawrence Erlbaum Associates.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence in favor of linguistic relativity. *Journal of Experimental Psychology: General, 129*, 369–398.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110–114.

Schmandt-Besserat, D. (1996). *How writing came about.* Austin: University of Texas Press.

Tallal, P. (2003). Language learning disabilities: Integrated research approaches. *Current Directions in Psychological Science, 12*, 206–211.

Trout, J. D. (2003). Biological specializations for speech: What can the animals tell us? *Current Directions in Psychological Science, 12*, 155–159.

Zuberbuhler, K. (2005). The phylogenetic roots of language: Evidence from primate communication and cognition. *Current Directions in Psychological Science, 14*, 126–130.

50

Thinking and Problem Solving

Kimberly Rynearson

Tarleton State University

Thinking is a broad term applied to the representation of concepts and procedures and the manipulation of information; acts such as reasoning, decision making, and problem solving are part of thinking. This chapter focuses on problem solving as one of the most complex intellectual acts. Humans must solve many problems per day. For example, the relatively simple act of driving to school or work requires us to solve problems related to the operation of a vehicle and the determination of the best route to our destination.

Researchers have studied the nature of problem solving for many years. This chapter focuses on the major theories and research findings generated from the last 100 years of problem-solving research. This chapter summarizes the methods of inquiry in this area of research in three parts: (a) behaviorist methods, (b) introspective methods, and (c) methods using computer modeling. Applications of these findings are discussed in the context of the development of expertise and the use of expert problem-solving strategies. Finally, the chapter concludes with a discussion of future directions for research on human problem solving.

DEFINITION OF PROBLEM SOLVING

Simply defined, problem solving is the act of determining and executing how to proceed from a given state to a desired goal state. Consider the common task of *driving to work or school.* To get from the given state, sitting in your dorm room, apartment, or home, to the desired goal state, sitting in class or in your office, requires you to solve problems ranging from the location of your keys to the physical execution of backing your car out of its parking space, observing traffic laws, choosing a route, negotiating traffic, and arriving at your destination in a timely manner. The first problem, locating your keys, can be broken down further into several cognitive acts: (a) perceiving that you need your keys to start the car, (b) generating ideas for where to search for your keys, (c) conducting a visual search for your keys, and (d) recognizing your keys when you see them. You can divide this relatively simple problem further into subproblems related to the retrieval of information from memory, the recognition of visual stimuli, and so forth. Thus, it is evident why psychologists consider problem solving one of the most complex intellectual acts. Although we solve many problems per day, the difficult and unfamiliar problems, such as how to complete the algebra homework assigned by a mathematics professor, make us notice how we solve problems; we tend to overlook the more mundane problem solving that makes up most of our day.

The following paragraphs present a historical overview of the theories that researchers have applied to the study of human problem solving. When reading these theories, notice that the theorists make different assumptions about human problem solving and use different techniques to isolate the most basic processes involved in this activity. I discuss their methodologies in more detail later.

THEORY AND MODELING

Gestalt Theory

Early in the 20th century, a group of German psychologists, most notably Max Wertheimer, Kurt Kofka, and Wolfgang Köhler, noted that our perception of environmental stimuli is marked by the tendency to organize what we see in predictable ways. Although the word "Gestalt" has no direct translation in English, it is loosely translated as "form," "shape," or "whole." The Gestalt theorists believed and demonstrated that humans tend to perceive stimuli as integrated, complete wholes. From this belief followed the assertion by the Gestalt theorists that the whole of an object is greater than the sum of all of its parts; something new or different is created when all the parts of a perception are assembled.

Perceptual Tendencies

Building on the basic premise that the whole is greater than the sum of its parts and the notion that we perceive visual stimuli in predictable ways, the Gestalt theorists proposed several perceptual tendencies that determine how we group elements of a stimulus's parts. These tendencies are (a) proximity—elements tend to be grouped together according to their nearness, (b) similarity—items similar in some respect or characteristic tend to be grouped together, (c) closure—items are grouped together if they tend to complete some known entity (i.e., a cube, face, etc.), and (d) simplicity—items are organized into simple figures according to symmetry, regularity, and smoothness. The perceptual tendencies, or principles of organization as the Gestalt theorists called them, relate to problem solving in that the perception of an individual stimulus or an entire problem is greater than the individual components of the problem. That is the reason that, when someone considers the *driving to work or school* task, the problem is different from the individual parts of the problem, down to the perception of a pile of metal as one's car keys. Essentially, the mind proceeds from understanding a problem in its discrete parts to recognizing the whole problem.

Insight

In addition to proposing the various perceptual tendencies that affect problem solving, research on animal problem solving conducted by Wolfgang Köhler provided support for the notion of *insight* in problem solving. In one of his most famous demonstrations of animal problem solving, Köhler (1927) studied the problem-solving ability of a chimpanzee named Sultan. From his cage, Sultan used a long pole to reach through the bars to grab and drag a bunch of bananas to him. When faced with a more difficult problem, two poles that individually were too short to reach the bananas outside of the cage, Sultan sulked and then suddenly put the poles together by placing one pole with a smaller diameter inside the other pole with a larger diameter. The resulting combined pole was long enough to reach the bananas. Köhler argued that Sultan's ability to solve this problem demonstrated *insight*, a sudden solution to a problem as the result of some type of discovery. True to Gestalt principles of organization, Köhler further argued that animals notice relations among objects, not necessarily the individual objects.

In summary, the principles of organization proposed by the Gestalt theorists provide some information about how animals and humans perceive problems and their component parts. However, the work of the Gestalt theorists marks the beginning of almost 100 years of research on problem solving. A more recent theory, problem space theory, defines the concept and proposes a simple framework for considering the act of problem solving.

Problem Space Theory

Similar to the Gestalt psychologists, Allan Newell and Herbert Simon (1972) emphasized the importance of the whole problem, which they called the *problem space.* The problem space, according to Newell and Simon, represents all possible configurations a problem can take. Problem solving begins with an *initial state*, the given situation and resources at hand, and ends with the *goal state*, a desired state of affairs generally associated with an optimal solution to a problem. To get from the initial state to the goal state, a problem solver must use a *set of operations*, tools or actions that change the person's current state and push that state closer to the goal state than the initial state. Conceptualized in this way, problem solving seems straightforward. However, Newell and Simon proposed that this relatively simple process has limitations, called *path constraints*. Path constraints essentially allow the problem solver to rule out some operators and methods to a problem solution.

Based on Newell and Simon's (1972) description of the problem space, we can describe problem solving as a search for the set of operations that moves a person most efficiently from the initial state to the goal state. The search for the set of operations is the heart of problem solving. When a problem is *well defined*, the problem space is identified clearly and the search for the best set of operations is easy. However, many of the problems we solve are *ill defined*; we do not recognize or retrieve from memory the initial state, goal state, or the set of operations. Ill-defined problems pose a challenge to the search for the set of operations. Several methods of selecting operators that researchers have proposed include conducting an exhaustive search or using various problem-solving heuristics.

Exhaustive Search

One logical choice for searching for the best set of operations is to engage in a systematic and exhaustive search of all possible solutions. Theorists call this type of search a *brute force search*. When considering how to drive to work or school from your home or apartment, you have a number of options for making your way out of your neighborhood, presumably to a thoroughfare that takes you closer to your destination. Some of those options are more or less direct routes to the thoroughfare. Regardless, they are options. A brute force search assumes that the problem solver considers every possible option. Thus, it is easy to see how a brute force search, which eventually allows the problem solver to find the best solution, can be a time-consuming process.

Heuristics

Heuristics offer another way to identify the set of operations and most efficient path through a problem space. Theorists loosely define heuristics as basic strategies, or rules of thumb, that often lead a problem solver to the correct solution. As you will see in the following paragraphs, some general problem-solving heuristics include the *hill-climbing strategy*, *means-ends analysis*, and *working backward.*

The *hill-climbing strategy* is what the name implies. Consider standing at the bottom of a hill; your goal state is to make it to the top of the hill. To ascend the hill, you look for the combination of footsteps (and ultimately a path) that takes you to the top of the hill in the most direct manner. Each step should move you closer to the goal. This last point also marks the most important constraint of this strategy. That is, to solve some problems, it is sometimes necessary to backtrack or to choose an operation that moves you away from the goal state. In such cases, you cannot use the hill-climbing strategy.

Students of psychology often equate the hill-climbing strategy to the problem of earning a college degree. A student begins at an initial state with no college credits. The goal, of course, is to complete all of the required credits for a particular degree program. Each semester or quarter, students complete one or more courses among those required for their degree programs, thereby making forward progress toward the goal state and graduation. Of course, one must acknowledge that in the past some students have not applied the hill-climbing strategy consistently to this particular problem. Changing majors, failing courses, and taking unnecessary electives are all examples of operations that do not meet the forward progress criterion of the hill-climbing strategy.

A second strategy that offers more flexibility in terms of executing the set of operations is known as *means-ends analysis*. Researchers characterize means-ends analysis as the evaluation of the problem space by comparing the initial state and goal state and attempting to identify ways to reduce the difference between these two states. When evaluating the differences between the initial state and the goal state, the problem solver often breaks the overall problem down into subproblems with subgoals. Solving the subproblems and meeting the subgoals allows problem solvers to reduce the overall distance between the initial state and the goal state so they become the same state.

We can apply means-ends analysis to the problem of earning a college degree. Most college degrees consist of a series of requirements that are grouped into courses that meet general education requirements, courses that meet the major requirements, courses that meet minor requirements, and courses that serve as electives. By breaking the college degree problem down into a series of subgoals, a student may choose to take courses to satisfy the requirements of each grouping in turn. That is, it is not uncommon for college students to take all courses for the general education requirements before beginning courses required for the major or minor. In summary, the assumptions of means-ends analysis allow this strategy to accommodate the need for a choice or an operation that appears to be a lateral move or a move away from the goal state, but one that eventually helps the problem solver meet a subgoal of the overall problem.

A third heuristic is known as *working backward*. Like the hill-climbing strategy, working backward is self-explanatory. A person solves a problem by working backward from the goal state to the initial state. Problem solvers use this heuristic most often when they know the goal state but not the initial state. An example of when to use such a strategy is when you check your bank balance at the end of the day. If you know all of the transactions that have posted during the day, you can work backward from your ending account balance to determine your account's beginning balance for the day.

Assuming that we can define problems in terms of a space with an initial state, a goal state, and a set of operations, a critical aspect of problem solving involves the search for the set of operations. Individuals search for the optimal set of operations to solve problems using a variety of methods. The choice between an exhaustive, brute force search and one of the heuristic strategies is constrained by the characteristics of the problem. However, there is another source of information that individuals bring to bear upon the problem solving process—background knowledge. Research on problem solving has proposed several ways that individuals use what they already know to devise solutions to problems. The following paragraphs describe how we use analogies to solve problems.

Theories of Analogical Reasoning

Occasionally when solving a problem, some aspect of the problem reminds us of another problem we have solved before. The previously solved problem can provide us with a model for how to solve the problem in front of us. That is, we draw an analogy between one problem and another. For example, you may solve the problem of driving to work or school regularly. When faced with the problem of driving to the grocery store, you may draw an analogy between previously solved problems and decide to rely on the strategies you used to determine how best to get from home to work or school when you decide how to get from home to the grocery store.

Using analogies to solve problems seems easy. However, most people fail to use them with any regularity. Research by Mary Gick and Keith Holyoak (1980) has suggested that analogies can assist problem solving, but that few individuals use them unless they are prompted to do so. Gick and Holyoak (1980) used Duncker's (1945) tumor problem to understand how individuals use analogies to solve problems:

> Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. To operate on the patient is impossible, but unless the tumor is destroyed, the patient will die. A kind of ray, at a sufficiently high intensity, can destroy the tumor. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At lower intensities the rays are harmless to healthy tissue, but will not affect the tumor. How can the rays be used to destroy the tumor without injuring the healthy tissue?

The solution to the problem is to divide the ray into several low-intensity rays so that no single ray will destroy healthy tissue. Then, the rays can be positioned at different positions around the body and focused on the tumor. Their combined effect will destroy the tumor without destroying the healthy tissue. Before reading the tumor problem, however, Gick and Holyoak asked one group of their participants to read a different problem, known as the attack dispersion story:

> A dictator ruled a small country from a fortress. The fortress was situated in the middle of the country and many roads radiated outward from it, like spokes on a wheel. A great general vowed to capture the fortress and free the country of the dictator. The general knew that if his entire army could attack the fortress at once it could be captured. But a spy reported that the dictator had planted mines on each of the roads. The mines were set so that small bodies of men could pass over them safely, since the dictator needed to be able to move troops and workers about; however, any large force would detonate the mines. Not only would this blow up the road, but the dictator would destroy many villages in retaliation. A full-scale direct attack on the fortress therefore seemed impossible.
>
> The general, however, was undaunted. He divided his army up into small groups and dispatched each group to the head of a different road. When all was ready he gave the signal, and each group charged down a different road. All of the small groups passed safely over the mines, and the army then attacked the fortress in full strength. In this way, the general was able to capture the fortress.

Gick and Holyoak found that roughly 76 percent of participants who read the attack dispersion story before reading Duncker's tumor problem drew the appropriate analogy and solved the tumor problem. When the researchers gave participants a hint that the attack dispersion solution may help them solve the tumor problem, 92 percent produced the correct solution. Thus, it seems that individuals may not use analogies to solve problems as readily as we may assume.

Researchers have proposed several theories to explain analogical reasoning. The most well-known theory, the *structure mapping theory*, proposes that the use of analogies depends on the mapping of elements from a source to a target (Gentner, 1983). The mapping process focuses not only on elements (objects) but also on relations among elements. According to structure mapping theory, individuals must search their background knowledge for sources that are similar to the target. Then, the individual must determine whether there is a good match between what is retrieved and the target. Problem solving by analogy, therefore, requires us to retrieve an old solution from memory and to recognize that it does or does not fit the current problem.

Keith Holyoak and Paul Thagard (1989) offered a different view of analogical reasoning. They proposed *multiconstraint theory*, which fits within structure mapping theory. Specifically, multiconstraint theory describes the factors that limit the analogies that problem solvers construct. Multiconstraint theory proposes three factors that govern how individuals use analogies. Specifically, the ease with which an individual uses an analogy to solve a problem depends upon similarities in structure, meaning, and purpose between the source problem and the target problem. Structural similarity refers to the extent that the source problem and target problem possess a similar structure. Similar structures allow problem solvers to map elements and relations among elements from the source problem to the target problem easily. Similarity in meaning further eases a problem solver's ability to perceive how the solution for one problem may apply to another. In Gick and Holyoak's (1980) research, for example, the fortress and troops were similar in meaning to the tumor and rays. Finally, multiconstraint theory's similarity of purpose refers to the extent that an individual recognizes the similarities between the goals in a source problem and a target problem. In the example of using an analogous solution for driving to work or school when trying to solve the problem of driving to the grocery store, there is high similarity in the purpose, structure, and meaning of the two problems.

In summary, researchers have proposed and applied several theories to the act of problem solving. The theories discussed here are not necessarily mutually exclusive. Rather, each can be applied to a different aspect of problem solving. Problem solving begins with the perception of the problem. According to the Gestalt theorists, we perceive problems like objects, as a whole. This tendency may affect our ability to represent accurately the initial state, goal state, and the set of operations needed to solve a problem. Our perceptual tendencies may further constrain our ability to choose a heuristic or recognize an appropriate analogous solution. Research in the area of problem solving has focused on discovering the conditions under which the assumptions of the theories presented here apply.

I describe the methodological traditions that researchers have used in the following section.

METHODS

Studying problem solving is a challenging process. When an individual generates a solution to a problem, sometimes it is easy to observe evidence of that success. For example, we can be sure that people have solved the problem of driving to work or school if they arrive at their destination in a timely manner. However, it is what happens between the initial state and goal state that has interested researchers and has proved more difficult to study. That is, researchers have used several methods for observing problem-solving processes (e.g., planning, retrieving problem-specific information from memory, choosing operators) that may be otherwise unobservable.

As in the description of Gestalt theory, early studies of problem solving focused on animal learning and problem solving. Although findings from animals may not always generalize to humans, researchers took notice of the structure of the experiments, the problems used, and the observations made. Although most real-life problem solving is quite complex and involves numerous subproblems, researchers sought to simplify problems so that they could isolate and observe various aspects of the problem-solving process.

Building on the methods used in animal research, a large portion of the problem-solving literature uses behaviorist research methods. Specifically, investigators provide research participants with structured problems and then observe their problem-solving behaviors, record the number of steps they use to solve a problem, and determine whether they generate the correct solution.

Investigators have often used the Tower of Hanoi problem to study problem solving. The Tower of Hanoi problem was invented by French mathematician Edouard Lucas in 1883. The most common version of this problem, the three-disk problem, requires the problem solver to move the disks from peg 1 (the initial state) to peg 3 (the goal state) given two conditions: (a) only one disk may be moved at a time, and (b) a larger disk may never be placed on top of a smaller disk (see Figure 50.1).

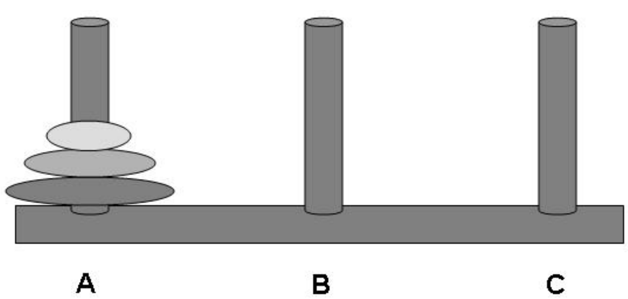

Figure 50.1 The Tower of Hanoi problem.

In one interesting study using the Tower of Hanoi, Fireman and Kose (2002) observed how successfully six- to eight-year-old children solved three-disk and four-disk versions of the Tower of Hanoi problem. Fireman and Kose divided their sample into three training groups and one control group. One training group practiced solving the three-disk problem, a second group received training in the correct solution, and a third group watched videotapes of their own previous attempts to solve the problem. Fireman and Kose observed the children's ability to solve the three-disk and four-disk problems immediately after training and one week later. The results showed that children who practiced independently and those who watched videotapes of their prior attempts made the most progress in their ability to solve the problems correctly. Additionally, their gains were maintained after one week. The methodology of this study and the Gick and Holyoak (1980) study described earlier are characteristic of most behavior-based studies of human problem solving in that the methods and outcomes of a participant's problem-solving efforts are recorded as evidence of unseen cognitive processes.

It is difficult to deny that problem solving is a cognitive act. Accordingly, some researchers have argued that to understand problem solving, you have to know what the individual is thinking while executing the behaviors that lead to a solution. To get this information, some of the research in problem solving has relied on introspection, or verbal "think-aloud" protocols generated by individuals while solving problems. The earliest use of introspection in psychology is attributed to Wilhelm Wundt in 1879. Later, Edward Titchener advocated the use of verbal protocols. More recently, Herbert Simon and Allen Newell defined the "think-aloud" protocol as a research methodology in which a research participant thinks aloud while engaging in some kind of cognitive activity.

The analysis of the think-aloud protocol is known as *verbal protocol analysis*. The primary advantage of this method is that it is the only way to know what information is in working memory during the problem-solving process. This advantage also relates to the primary limitation of this methodology. That is, individuals verbalize only thoughts and processes of which they are consciously aware. Automatic processes, which are executed outside of conscious awareness, may not be articulated in verbal protocols. Despite this limitation, the use of think-aloud protocols continues, particularly in the field of education. Instructors often ask learners to verbalize what they are thinking as they solve mathematics problems or as they read texts.

A final research methodology used in the study of problem solving involves the use of computers to model the problem-solving behavior of humans. Researchers build computer programs that mimic the strategies that human problem solvers often use to solve a problem. Then, the researchers compare the computer's attempts to solve the problem to the existing data on how humans solve the same problem. The primary benefit of this particular research methodology is that building a computer model that fits data from human problem solvers requires the researcher to think carefully and to define every step in the problem-solving process explicitly.

The most well-known example of such a computer model is Newell and Simon's (1972) *General Problem Solver*. Newell and Simon compared the verbal protocols from humans who attempted to solve abstract logic problems to the steps generated by the General Problem Solver and found a high degree of similarity between the self-reported problem-solving actions of humans and the General Problem Solver. Additionally, their research noted that humans and the computer model relied on means-ends analysis to solve the logic problems. The humans and the model created subgoals to reduce the difference between the initial state and goal state of the problem space. Since Newell and Simon's work, researchers have used computer modeling to study how people solve everyday problems such as reading single words, playing games, making medical diagnoses, and even reasoning about a particular law's application to a case.

In conclusion, research on problem solving has relied on three methodologies, direct observation of problem-solving behaviors, think-aloud protocols, and computer modeling. For theory building, computer modeling seems to be the most promising methodology. As already stated, a computer model that effectively accounts for the behavioral or think-aloud data requires the researcher to state the specific assumptions and mechanisms required to solve various problems. This level of specificity creates opportunities for researchers to test the assumptions of theories of human problem solving generally and within specific domains (e.g., mathematics, reading).

APPLICATIONS

Problem solving is an everyday occurrence. Thus, the opportunities for application of the concepts presented in this chapter are numerous. This section discusses the practical applications of what psychologists have learned from research on human and animal problem solving. Specifically, the research on problem solving suggests some specific strategies an individual may use to avoid common mistakes in problem solving. Finally, this section describes how experts solve problems in their domain of expertise.

Five Rules for Problem Solving

Henry Ellis (1978) identified five basic rules a person can apply to any problem-solving situation. Consider how each of these rules relates to the theory and research that we have considered in this chapter. For each of Ellis's rules, try to generate your own example of when to apply each rule.

1. *Understand the Problem*

The first step to solving any problem is to understand the current problem. As discussed earlier in this chapter, most real-life problems are ill structured, meaning that the initial state, goal state, and the set of operations (or possibly all three) are not known or are unclear. Thus, successful problem solving starts with a clear understanding of the problem space. Thus, it is best to begin solving a problem with a thorough analysis of the problem to ensure understanding of the problem space.

2. *Remember the Problem*

A second step to solving problems successfully is to remember the problem and the goal throughout the problem-solving process. Often, we lose sight of the goal and instead solve a different problem, not the problem before us. A common example is forgetting a question someone asks you and instead answering a completely different question.

3. *Identify Alternative Hypotheses*

When problem solving, it is good practice to generate multiple hypotheses, or guesses about the set of operations, for the best strategy or solution to the problem. We often make a mistake when we settle upon a solution before considering all relevant and reasonable alternatives.

4. *Acquire Coping Strategies*

Because life presents many challenges, the problems we solve daily can be difficult. Frustration is a barrier to problem-solving success, as are *functional fixedness* and *response set*. Functional fixedness is a tendency to be somewhat rigid in how we think about an object's function. This rigidity often prevents us from recognizing novel uses of objects that may help us solve a problem. For example, our failure to recognize a butter knife's ability to do the job of a flat-head screwdriver is an example of functional fixedness. Response set, however, relates to inflexibility in the problem-solving process. Response set is our tendency to apply the same set of operations to problems that have similar initial states and goal states. However, this strategy breaks down when some aspect of the problem space changes, making the set of operations that were successful for previous problems the wrong choice. Essentially, we become set in our response to a problem. This tendency prevents us from noticing slight differences in problems that may suggest a simpler or different solution process. Thus, it is important to be aware of these common barriers to problem solving and to take steps to avoid them.

5. *Evaluate the Final Hypothesis*

In everyday problem solving, it is important to evaluate the practicality of your solutions. That is, the shortest route from your home or apartment to work or school may also be the route that is most likely to be congested with traffic when you will be traveling. Thus, the more practical solution may be a less direct route that has less traffic. The point is that you should evaluate your final hypothesis for a solution to determine if it is the best or only solution to the problem, given the constraints of the problem space.

Expert Problem Solving

In addition to applying any or all of the five rules mentioned, try approaching a problem-solving situation the way an expert does. An expert is someone who is highly informed on a particular subject and is very good at solving problems in that domain. Consider Tiger Woods. Mr. Woods is an expert golfer. He knows more about golf than the average individual does, and he is very good at figuring out how to solve the most obvious golf-related problem, how to get the ball from the tee to the hole in the fewest shots. What has research uncovered about expert problem solvers?

First, experts simply know more about their domain of expertise than novices do. Researchers generally recognize the *10-year rule*, the idea that individuals need 10 years of experience or 10,000 hours of practice in a domain to achieve expertise (for a review of research on expertise, see Ericcson, Charness, Hoffman, & Feltovich, 2006). Given this rule, it is conceivable that an expert has more opportunities to learn about his or her area of expertise and more opportunities to store that information in memory. Research supports this idea. William Chase and Herbert Simon (1973) reported that chess masters have superior memory, in comparison to chess novices, for the midgame position of chess pieces on a chessboard, even when exposed to the board briefly. When the pieces are placed randomly on the board, however, chess masters demonstrate no advantage over chess novices for memory of the pieces' positions. These findings suggest that chess masters, due to experience, have stored many legal board positions in memory. Thus, experts simply have more experience in their domain and know more about it than do novices.

Second, experts organize information in memory differently from novices. Due to their additional knowledge and practice in a domain, experts have rich, interconnected representations of information in their area of expertise. Micheline Chi and her colleagues have compared physics masters to physics novices to observe differences in how the two groups represent problems. Chi, Feltovich, and Glaser (1981) noted that novices and experts, when presented with a variety of physics problems, showed differences in how they grouped the problems. Novices tended to focus on physical similarities of the problems, for example, by grouping all problems related to inclined planes together. Experts, however, grouped problems according to the underlying physical concepts (e.g., conservation of energy) related to each problem. These and other related findings suggest that experts understand their content domain at a deeper level and, therefore, organize information differently.

Finally, experts move through the problem space differently from novices. Experts are less likely to use the working backward strategy when solving problems in their domain of expertise. Instead, they work forward from the initial state to the goal state (Larkin, McDermott, D. Simon, & H. Simon, 1980). Novices, however, tend to work backward from the goal to the initial state. Jose Arocha and Vimla Patel (1995) compared diagnostic reasoning of beginning, intermediate, and advanced medical students. Rather than working forward from the symptoms to a diagnosis, the students tended to use a backward search from the diagnosis to the symptoms. In this study, initial case information suggested diagnosis of a common disease. The least experienced novices in the study ignored subsequent case information that would have required them to change their hypotheses about the diagnosis. Therefore, experts use their well-organized knowledge in a domain to recognize a problem and solve it by moving continually toward a solution.

FUTURE DIRECTIONS

Although the field of problem solving, particularly expert problem solving, has been researched extensively over the past 30 years, there are related theories to be tested and other measurement techniques to apply. John Anderson's (1983, 1990, 1993) ACT* (*adaptive control of thought*) model (and later ACT-R, adaptive control of thought—rational) is a model of knowledge representation that proposes a cognitive architecture for basic cognitive operations. ACT-R describes human knowledge as having of two kinds of memory: *declarative memory*, which consists of an interconnected series of nodes representing information, and *procedural (or production) memory,* which comprises series of productions, or *if-then* pairs, that relate to various conditions and their associated actions. Declarative memory and procedural memory interact with working memory. In Anderson's model, working memory receives information from the outside world and stores it in declarative memory or retrieves old information from declarative memory for current use. Working memory also interacts with procedural memory in that working memory seeks to match the *if* conditions so that the *then* actions can be executed. Although Anderson's model is more complex than what is summarized here, the application of this model to problem solving is clear. Problem solving, according to this model, likely takes place in two stages, an *interpretative stage* and a *procedural stage*. When individuals encounter a problem, they retrieve problem-solving examples from declarative memory. This is the interpretative stage. The procedural stage follows and involves the evaluation and execution of related if-then (condition-action) pairs. Research on Anderson's model suggests that problem solving proceeds in a manner similar to means-ends analysis. That is, there is constant reevaluation of the problem space and subsequent execution of steps to reduce the difference between the current state and the goal state. In summary, Anderson's cognitive architecture holds promise for further application to the study of human problem solving. Specifically, researchers have applied the ACT-R model to measurements of human problem-solving performance generated by brain-imaging technologies.

Although researchers generally study human cognition by observing behaviors that are indicative of various cognitive acts (i.e., word recall as evidence of memory), researchers recently have turned to brain-imaging technologies to gain a further understanding of what happens in the brain when we try to recognize a face, retrieve a word, make a decision, or solve a problem. Functional magnetic resonance imaging (fMRI) is one relatively new technique that measures activity in the brain. Simply put, fMRI measures blood flow within the brain that occurs in response to neural activity. Research using fMRI indicates that blood flow in several distinct areas of the brain, namely the left Dorsolateral Prefrontal Cortex (DLPFC) and the Anterior Cingulate Cortex (ACC), appear to be related to human problem solving (see Cazalis et al., 2006). In a recent study of problem solving, Anderson, Albert, and Fincham (2005) used fMRI analysis to track patterns of brain activity predicted by the ACT-R cognitive architecture. In this study, fMRI confirmed brain activity in the parietal region related to planning moves to solve the Tower of Hanoi problem. Thus, Anderson et al.'s study provides evidence that investigators can use fMRI to examine human problem solving and to verify the assumptions of cognitive theories and models.

SUMMARY

This chapter has summarized the major theories, research findings, and applications of research in the field of human problem solving. As one of the most complex acts of human cognition, the study of problem solving will continue, with greater emphasis on the generation of computer-based models that predict data from behavioral observations and data generated from imaging studies of neural activity within the brain.

REFERENCES AND FURTHER READINGS

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35–44.

Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. *Journal of Cognitive Neuroscience, 17*(8), 1261–1274.

Arocha, J. F., & Patel, V. L. (1995). Novice diagnostic reasoning in medicine: Accounting for clinical evidence. *Journal of the Learning Sciences, 4*, 355–384.

Cazalis, F., Feydy, A., Valabregue, R., Pelegrini-Issac, M., Pierot, L., & Azouvi, P. (2006). fMRI study of problem-solving after severe traumatic brain injury. *Brain Injury, 20*(10), 1019–1028.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Duncker, K. (1945). On problem-solving. *Psychological Monographs: General and Applied, 58*(5, Whole No. 270).

Ellis, H. C. (1978). *Fundamentals of human learning, memory, and cognition.* Dubuque, IA: Wm. C. Brown.

Ericcson, K. A., Charness, N., Hoffman, R. R., & Feltovich, P. J. (Eds.). (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge, UK: Cambridge University Press.

Fireman, G., & Kose, G. (2002). The effect of self-observation on children's problem solving. *Journal of Genetic Psychology, 164*(4), 410–423.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 115–170.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355.

Hayes, J. (1985). Three problems in teaching general skills. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning* skills (pp. 391–406). Hillsdale, NJ: Erlbaum.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295–355.

Köhler, W. (1927). *The mentality of apes.* New York: Harcourt Brace.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335–1342.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Plaut, D. C. (1999). Computational modeling of word reading, acquired dyslexia, and remediation. In R. M. Klein & P. A. McMullen (Eds.), *Converging methods in reading and dyslexia* (pp. 339–372). Cambridge, MA: MIT Press.

51

Critical Thinking

Diane F. Halpern
Claremont McKenna College

Patrick Williams
Claremont Graduate University

The National Center on Education and the Economy (2007) issued a report in which they envisioned the needs of the future workforce. They provide an educated prediction about how we will work and, as a consequence, how we will live during the next several decades. If you are embarking on your career or just planning on being alive for several more decades, it is important reading. The report, titled "Tough Choices or Tough Times," depicts the prototypical industry in the next 10 years (if all goes well!) as resting on a base of routine work that is done by both people and machines, topped with creative work where critical thinking skills are essential. Some examples of high-level work, located at the peak of the pyramid-shaped depiction of the world of work, are research, development, design, marketing and sales, and global supply chain management. This list of high-level jobs that require critical thinking brings to mind a key question that was posed by Earl Hunt, a cognitive psychologist at the University of Washington. In his book on the future of the U.S. workforce, Hunt (1995) asked, quite simply, "Will We Be Smart Enough?" The answer to that question has significant repercussions for our society and for young people preparing for their future at a time in history when the ability to engage in high-level thinking skills will determine the quality of your work life, your income level, and the role you will play in the rapidly developing and technologically savvy world.

The good news is that creative and critical thinking skills can be taught and learned. All we need to do now in order to remain competitive (and we add cooperative) as a nation is persuade those people with the power to set curricular agendas and national imperatives to invest time, money, and energy into making critical-thinking instruction a national priority. Instruction in critical thinking can begin in elementary school and continue through graduate and professional education, with age-appropriate learning objectives and examples. The research literature on high-order learning has made it clear that both high- and low-achieving students benefit from explicit training and repeated practice in critical thinking. There are ample data to support our claim that critical-thinking skills can be learned and transferred to novel situations when students receive explicit instruction designed to foster transfer (Bangert-Drowns & Bankert, 1990; Halpern, 2003).

Instruction in critical thinking is for everybody. No society can prepare its workforce for the future if there is an underlying belief that the work that requires critical thought should be reserved for the educated elite. In our increasingly complex and technological world, critical-thinking skills are needed by a growing portion of the population. Manual labor is increasingly done by machines, and repetitious and unskilled jobs offer little or no opportunity for personal or financial growth. According to Browne and Keeley (2004), the rapid acceleration of the complexity of the world drives the need for more and better education in critical thinking and other high-level skills such as oral communication, writing, and technology. But what exactly is critical thinking, and how does one learn to think critically?

DEFINING CRITICAL THINKING

Jones and colleagues (1995) surveyed 500 policy makers, employers, and educators and found a consensus for the meaning of critical thinking. It is a broad term described as reasoning in an open-ended manner that includes an unlimited number of solutions. It also involves constructing a reasoned argument with evidence and reasoning that supports the conclusion. Critical thinking should be self-directed and "fair-minded," with clarity about the nature of a problem, the way generalizations are made, the way evidence is presented, and how conclusions are drawn.

Halpern (2003) proposed one of the most comprehensive definitions of critical thinking. She defined critical thinking as the appropriate use of cognitive skills that increase the probability of a desirable outcome. This definition includes the idea that critical thinking is not just "fair" or "reflective"; it also includes thinking in ways that are related to better outcomes. Halpern's definition also entails the idea that critical thinking is purposeful, reasoned, and goal-directed. In other words, when people are engaged in critical thinking they are attempting to move toward a desired goal or outcome. It is the sort of thinking that is used when solving problems, formulating inferences, calculating likelihoods, and making decisions. Critical thinking also involves the conscious evaluation of thought processes used to achieve a desired goal (e.g., make a decision) as well as the outcome of those processes (e.g., how good the decision was or how well the problem was solved).

The act of evaluating one's own thought processes is known as *metacognition*. More colloquially, metacognition is thinking about one's own thinking. Costa (2001) offered a simple example to illustrate the usual findings that too often most people do not reflect on the "how" of their own thinking. He asked a large group of people "How much is one half of two plus two?" Quickly—what is your answer? How did you think about this question? Did you mentally take half of two and then add it to two, or did you sum two plus two and then take half? It is likely that you did not think about the way in which you solved this problem or even realized that alternative solution strategies would produce different answers. Costa's point is that many people fail to think about the solution strategies they use and instead apply a ready-known strategy that is easily accessible in memory.

With proper training in critical thinking, people can understand better the multiple challenges of their environment and society at large. Some skills involved in critical thinking include, for example, evaluating the cause of an event or events; giving reasons that support a conclusion; recognizing and criticizing assumptions that are either implicit or explicit in an argument; and developing an orderly plan for comprehension of a complex topic such as prioritizing the order in which "pieces" of evidence or arguments are examined and rated for quality.

The use of analogy in problem solving is an example of a critical-thinking skill that often provides novel or creative responses. A concrete analogy can make an abstract concept easier to comprehend and remember. Halpern's definition allows for creative and critical thinking to exist under the same rubric because creativity is conceptualized as a novel response that increases the probability of a desirable outcome. Thus, creative thinking is one type of critical thinking. Its sole distinction from other types of critical thinking is that the response is novel. In addition to creative thinking, a critical thinker should also understand basic research principles and be able to present a coherent and persuasive argument about a controversial topic. It is not a coincidence that these outcomes are the same outcomes that many colleges and universities list as desired outcomes for their graduates.

Assumptions for Critical-Thinking Instruction

The idea that critical-thinking skills can be taught is predicated on two assumptions: (a) Thinking skills are clearly definable and identifiable and can be taught and applied appropriately, and (b) students will be more effective thinkers if thinking skills are recognized and applied. It is clear that students can learn critical-thinking skills from both explicit classes on critical thinking as well as from critical-thinking skills instruction that is embedded within the curriculum of a class. The advantage of embedding critical thinking into course work is that this instructional approach does not require additional courses in an already over packed curriculum.

The disadvantage of this approach is that it is heavily reliant on the instructor in content-area classes (e.g., literature, history, psychology) to provide instruction in critical thinking in addition to "covering" the basic content of the course. Most instructors do not have the necessary background knowledge to teach broad critical-thinking skills in ways that would transfer across academic disciplines. As experts in a content area, teachers tend to focus instruction on the topic they are expected to teach. In fact, it would be difficult to argue that an emphasis on a teacher's content area of expertise is not appropriate, given that is the way our curriculum is arranged. The problem with relying on imbedded instruction for critical-thinking skills is that the skills may never be taught and even those teachers who use critical thinking for understanding their content matter may not provide the necessary instruction in how to think critically.

When students learn critical-thinking skills within a specific discipline, those skills are less likely to transfer to other academic disciplines, or more important, to the real world. Sternberg (1997) recommended that to increase the chances of transferring critical-thinking skills from the classroom to everyday situations, instructors should use real-life issues when teaching problem-solving skills. Halpern advocated for critical-thinking instruction that uses multiple examples from different disciplines, so that students learn to focus on the critical-thinking skill that is being taught and so that they learn when it is appropriate to use that skill in many different contexts. Along similar lines, Staib (2003) suggested the use of real-life role-play, case studies, discussion, and student-instructor interaction as effective tools in critical-thinking instruction.

Although some elementary and high school teachers are familiar with critical thinking as its own content area and will teach critical thinking along with their usual course topics, it is most often the case that formal education in critical thinking does not begin until postsecondary education. A critical-thinking course is often included in undergraduate education, with opportunities to practice those skills in courses throughout the curriculum. If critical-thinking instruction occurs only at the college level, then many students, particularly those from poor or disadvantaged populations with low college-going rates, will not have the opportunity to learn the important critical-thinking skills they will need for their future employment and for managing the complexities of everyday life, such as determining which phone plan is best for a particular person or how to decide about the myriad complex issues that face them as voters. The inclusion of critical-thinking instruction is most urgent in high schools, surpassing the need in both middle and elementary schools. What high school students generally lack upon graduation, diploma or not, are the tools to deal with higher education and the real world. Thus, we are advocating for courses designed to teach critical-thinking skills in ways that facilitate transfer in all high schools and colleges.

THEORETICAL PERSPECTIVES

There are some naysayers who contend that the emphasis on critical-thinking instruction is just another educational fad that will go away if they wait long enough. It is likely to be a very long wait. The new fad of critical thinking can be traced back to the famous American educator, John Dewey. Dewey's writing has been influential since the very early 1900s and continues to be so today. He advocated for a progressive education in the instruction of both skeptical and reflective inquiry. Reflection, skepticism, and inquiry are important components of what is now referred to as critical thinking. When students are taught how to solve problems, they learn that they must consider multiple types of evidence to understand the problem, which includes context, alternative assumptions, reasons, explanations, motives, and costs and benefits for alternative solutions. They must be able to tolerate ambiguity while they are working to reduce uncertainty (for example, gathering additional information). These basic ideas may be found in Dewey's many books on learning (one such book, *Democracy and Education,* published in 1916, is available free online at http://www.ilt.columbia.edu/Publications/dewey.html). He thought of schools as a sort of repair organ for society. Through what is now called critical thinking, schools can refocus on ways to provide students with necessary thinking skills so that they can live up to Dewey's legacy and ultimately correct modern societal ailments. In this way, progressive education can lead to a society that can improve itself.

Halpern (1998) proposed a four-part model for teaching thinking skills so they will transfer across domains of knowledge. The component parts include (a) a dispositional or attitudinal component, (b) instruction in and practice with critical-thinking skills, (c) structure-training activities designed to facilitate transfer across contexts, and (d) a metacognitive component used to direct and assess thinking. Each of these components is grounded in theory and research in cognitive psychology.

The dispositional or attitudinal component of critical thinking is distinct from any of the performance or contemplative components. Some people have excellent critical-thinking skills and may recognize when the skills are needed, but they also may choose not to engage in the effortful process of using them. This situation exemplifies the distinction between what people can do and what they actually do in real-world contexts. People with dispositions that are not congruent with doing the difficult work of critical thinking appear content with giving "cursory" thought to complex issues. Such people are often sanguine when decisions have poor outcomes, resolving their low level of effort in the decision-making process and the poor outcome it produced by assuming that the additional effort of thinking critically would not have helped or that there is little they can do to increase the probability of desirable outcome. For example, consider a student who has to decide where to live. Instead of putting in the hard work of gathering information about alternatives and evaluating them, she picks an apartment that is too expensive, too far from transportation, or unsafe. When any of these problems becomes apparent (after moving in), she might justify the failure to use critical-thinking skills in her decision-making process by stating that she had few choices (but she never checked out other apartments) or that it is just impossible to live on the money she has (when she never considered less expensive living arrangements).

Instruction and practice in critical thinking are essential because learners need to know what critical thinking skills are available, how to use them, and to how to select the right skill for any given situation. For example, when listening to a political candidate speak about all of the great accomplishments that will follow from her election, it is not an automatic response to think about the cost of what has been promised, what won't happen if we are paying for one desirable program and not funding others, or how the promised achievements will be accomplished. With practice in a variety of domains—the sciences, humanities, social sciences, and the popular media, for example—critical thinkers learn to recognize which critical thinking skills are appropriate for the situation.

The third part of Halpern's (1998) four-part model focuses on structure training. Linguists talk about surface and deep structure when describing language. Deep structure is the meaning part of language; surface structure refers to the words used to convey the meaning. This distinction is useful in thinking about structure-training in critical thinking. Suppose that a student learns about the need to consider base rates when making judgments. If this critical-thinking skill is taught so that the learner understands that there is a need to think about base rates when making a judgment about a sample taken from a population with known characteristics, then he should be able to use this skill with a diverse array of topics that have this underlying structure, such as thinking about the number of people in a sample who are likely to be aggressive or likely to be Democrats (or both aggressive and Democrats) or the likelihood that your neighbor's dog has fleas or rabies (or both). He will be able to see beyond the surface or topical level of the issue being considered to discern the structural similarities in these examples and consider ways that the sample might not be representative of the population from which it was drawn.

The fourth component in Halpern's (1998) model is metacognition. In the example just provided, in which a learner sees that the need to think about base rates cuts across contexts as diverse as aggressive people and dogs with fleas, he will consciously consider how to make those judgments, and after considering base rates, he will consider if the answer is likely to be correct. Critical thinkers will also consider whether there are other skills that might be useful (e.g., is there anything in the context or the sample that might alter his decision to use base rates). Conscious reflections on the process that went into deciding on an answer and the quality of the answer are necessary components of being a critical thinker.

Sternberg (1997) offered a tripartite theory of successful thinking. His three broad groupings of thinking skills include analytical thinking skills, creative thinking skills, and practical thinking skills. According to Sternberg, students generally have a preference for one type of thinking skill over the other two, and there is evidence to show that when students are taught predominantly with the method that matches their preferred thinking skill, learning is increased. Sternberg recognized the problem with advocating that students be taught in three different ways, so ideally, these three thinking skills should work together to form what he calls "successful intelligence." Analytical thinking skills are most valued in the school settings, the most heavily weighted on traditional standardized tests, and include analyzing, critiquing, judging, evaluating, comparing, contrasting, and assessing. Creative thinking skills include creating, discovering, inventing, imagining, supposing, and hypothesizing. Last, practical thinking skills include applying, using, and practicing the other thinking skills. Creative and practical skills are, for the most part, independent of the kind of skills that are assessed in traditional measures of intelligence.

In thinking critically, students should be able to evaluate what their preferred thinking skills are through metacognition. Metacognition is the ability to engage in self-correction through reflection upon thoughts and thought processes. Metacognition can be thought of as the executive "boss" function that acts as a guide to how someone uses different strategies and makes decisions on how best to allocate limited cognitive resources according to task demands. More simply, metacognition is the personal knowledge and understanding of one's cognitive processes. Given this definition, a metacognitive understanding of their intellectual strengths and weaknesses should allow people to modify and strengthen weaknesses in their thinking while utilizing their own particularly strong thinking skills. In this way, a critical thinker can learn from mistakes because he or she is able to identify mistakes, which makes improvement more likely to occur.

Kruger and Dunning (1999) provided a good example of the power of metacognition. They showed that people who scored very low in ability within a particular content area (participants scored in the 12th percentile) judged themselves to be significantly higher (in the 62nd percentile). This research provides evidence that when people have little knowledge of a content area, they tend to consider themselves much more knowledgeable than they are. The implications of these findings are that some people will never learn from their mistakes because they simply cannot recognize the mistakes that they make. The inability to see one's own errors is a clear barrier to improvement. Knowledge of competence in a given area, through metacognition, then leads to improvement of skill in that area.

METHODS OF ASSESSING CRITICAL THINKING

A strong body of evidence exists to support the conclusion that specific instruction in critical-thinking skills will enhance critical thinking when the instruction includes a wide range of contexts. If critical thinking is taught for just a few select contexts, it is less likely that those skills will be learned at a deep level that will generalize to novel real-life situations. In creating a critical-thinking assessment that can determine if critical-thinking skills transfer to everyday situations, Halpern (2001) created an online assessment of critical thinking in which problems are posed in the context of everyday situations (http://berger.cmc.edu/rodinproject). The Halpern critical-thinking assessment covers five broad categories of critical-thinking skills: verbal reasoning, argument analysis, hypothesis testing, likelihood and probability, and decision making and problem solving. It is a unique assessment in that each question begins with a common, everyday scenario. The first part of each question requires a constructed (written) response to a question about the scenario. The constructed response is followed by a similar question that is presented with a multiple-choice or multiple-rating format. The distinction between these two response types is based on the cognitive distinction between generating a response and recognizing a correct response. Everyone is more likely to respond correctly to any question when alternative answers are presented in a multiple-choice format than when the participant is required to construct a response without the benefit of seeing possible alternative answers. With two types of response formats, the test requires both the spontaneous application of critical thinking skills through constructed responses and the recognition of appropriate answers through a multiple-choice response format.

The Halpern critical-thinking assessment is ecologically valid, meaning that the scenarios are representative of examples found in newspapers and everyday discussions. The following is an example question from the online assessment that requires the construction of an appropriate response:

> A recent report in a magazine for parents showed that adolescents who smoke cigarettes also tend to get low grades in school. As the number of cigarettes smoked each day increased, GPA decreased. One suggestion made in this report was that we could improve school achievement by preventing adolescents from smoking.
> —Based on this information, would you support this idea as a way of improving the school achievement of adolescents who smoke?
> —Type "yes" or "no" and explain why or why not.

After writing a response, the same question could be posed followed by a selection of responses of varying degrees of suitability. These responses could either be ranked by how appropriate they are or the one best choice could be chosen.

Nisbett (1993) and his coauthors compiled a book of readings in which each chapter is a different study of success in teaching and learning critical-thinking skills. In one of the studies he described, researchers phoned the homes of students after these students had completed a course that taught about regression to the mean—a difficult concept for many people to learn. The investigators phoned the students under the guise of conducting a survey. They posed questions that required students to apply their recently learned knowledge about regression to the mean to a novel example while they were in their home environment. The researchers found that the students spontaneously applied the critical-thinking principles they had learned in class. This finding is particularly significant because school-related cues about the appropriate responses were absent. In addition to this study, Nisbett's edited book includes 16 chapters showing that rules of logic, statistics, causal deduction, and cost-benefit analysis can be taught in ways that will generalize to a variety of settings. He also found that logical and statistical errors, two common errors of inference, could be substantially reduced with only very brief training. Other independent reviewers came to a similar conclusion about the effects of courses designed to promote critical thinking. These examples demonstrate that it is not always difficult to help adults think more critically in ways that have long-lasting and cross-contextual effects.

Several studies of high schools found that most do not properly train students to deal with the real-world demands of work and higher education. Thus, an increased emphasis on critical-thinking instruction is important if American schools are to prepare their students for the rapidly changing world in which they will spend their adult life. For example, a 2005 report by the National Governor's Association found that of 10,000 high school students, over one third rated their critical-thinking skills as fair-to-poor. Although almost everyone agreed that critical-thinking instruction is important at all educational levels, 83 percent of respondents to a national survey expressed that the need for reform is greatest at the high school level.

Moseley, Baumfield, and Elliott (2006) reviewed the vast literature on critical-thinking instruction and found that although most educators agree about the importance of critical-thinking instruction, they do not agree on the best method to teach it. Both explicit training of critical-thinking skills and critical-thinking skills training that is imbedded into already existing curriculum have been shown to lead to increases in critical thinking. However, there is robust evidence that explicit instruction of critical-thinking skills is the preferred method for high school students.

Current research by Marin-Burkhart and Halpern (unpublished) supports the hypothesis that explicit instruction in thinking strategies is more effective than imbedded instructions in that explicit instruction is more likely to lead to transfer of critical-thinking skills to challenges faced in everyday situations. They found significant gains from only a few weeks of instruction, so courses need not be lengthy as long as instruction is taught for transfer and there are opportunities to practice what is learned in their regular classes. They found that students can learn critical-thinking skills in a class designed as an afterschool program or during the day without seriously disrupting normal curricular operations. In a meta-analysis of explicit critical thinking skills, Bangert-Drowns and Bankert (1990) found that an intensive, practical skills orientation was most effective in generalizing skills to everyday situations.

APPLICATIONS

Regarding the applications of critical thinking outside of the classroom, Moseley et al.'s (2006) review of 23 highly relevant studies whittled down from 6,500 chapters, articles, and papers found no negative effects of critical-thinking instruction on a range of noncurriculum measures. When students make the effort not only to learn critical-thinking skills but also to adopt a critical-thinking frame of mind, they learn to apply critical thinking to many everyday situations. A frame of mind for critical thinking requires a concern for accuracy and persistence, as opposed to what is familiar and comfortable. Often the hardest part of being a critical thinker is staying open to new ideas and suppressing immediate and easy responses.

COMPARISONS

Critical-thinking instruction is sometimes likened to courses that aim to increase intelligence because it seems logical to many people that if one becomes a better critical thinker, then intelligence must also increase. Much like intelligence, there has been disagreement as to whether critical thinking can be taught or improved throughout life, but unlike intelligence, few people believe that the ability to think critically is an inborn or innate skill that is relatively immune to educational practices or environmental effects. Fortunately, critical thinking does not carry all of the negative baggage associated with the concept of intelligence, thus, although they are similar concepts, there are also important differences in how people think about these related constructs.

Clearly, critical-thinking instruction should include a metacognitive function. Learning about learning and memory has a beneficial effect on all cognitive processes. If an instructor, for example, models his or her own thinking by thinking aloud or through other symbolic systems, a student can do the same until such habits become conscious, automatic, and most important, available for improvement.

FUTURE DIRECTIONS

Critical-thinking skills are increasingly included with traditional skills such as mathematics, writing, and oral communication as fundamental goals of education. It is interesting to note that very few people ever ask if writing skills transfer beyond writing courses or if mathematical skills transfer beyond mathematics courses, yet this question is often raised about critical-thinking courses, and we expect that much future research will examine the questions of how to make transfer more likely and how to assess critical thinking in natural contexts. Nothing in learning theory suggests that critical-thinking skills could not be learned or that they would not transfer, but these seem to be persisting questions in the literature. There is also little debate about the value of improving critical thinking. In the real world beyond the classroom, a lack of critical-thinking skills leaves one vulnerable and disempowered. To access resources in education and job training, the ability to think critically will only become increasingly important. Traditionally, kindergarten through twelfth grade curriculum has focused on pure didacticism, or memorization of content information. Too often, memorization is also stressed in typical standards-based assessments. To prepare future generations for an increasingly uncertain world, more focus must be shifted away from didacticism and toward more active participation by students in the learning process.

Although educators (almost) universally accept critical-thinking skills as an essential and central element of effective instruction, the task of fostering critical thinking is a daunting one. Educational standards at the state and national level describe critical thinking as integral to the learning process, yet it is largely recall for which students are tested on standards-based assessments. To foster greater critical thinking in the future, resources should be better allocated on teacher training to facilitate learning and applying and assessing critical thinking instead of primarily on student testing and on teacher assessment. A new approach would recognize the need for accountability and the important role of assessment while simultaneously recognizing the need to engage students in the learning process.

In traditional classrooms, students are treated as passive receptacles in which an instructor places knowledge for later recall. Contemporary research in cognitive psychology recognizes that learning is a constructive process whereby information is transmitted in pieces from an instructor to students; students must actively construct concepts, with the instructor acting as a cognitive guide. Constructivism has been influential in recent curriculum design to foster critical thinking. This theory of learning takes aim at educational tradition that treats learning as a one-way hierarchy, starting with content knowledge, moving through comprehension and application, and only in the end encompassing critical-thinking skills such as analysis, synthesis, and evaluation. As education continues to be reformed using knowledge about the way the mind works and curricular content continues to shift toward the goal of enhancing critical thinking, instruction in critical thinking will be regarded as an essential element of a whole education to prepare students for the problems faced in everyday situations.

SUMMARY

In fields as diverse as ecology and political science, agriculture and warfare, biology and human development, our future as a country and the future of our planet depend on an educated populace who can think critically about issues of local and global concern. There is a large body of evidence showing that critical thinking can be taught and learned in ways that transfer across settings and time. Although there are multiple ways to help students improve their critical-thinking abilities, there is strong support for direct and conscious instruction in critical-thinking skills, using a wide variety of examples and helping learners think about the thinking process. The need for critical-thinking instruction at all levels of education is a fundamental concern because only a critically thinking workforce and citizenry can successfully tackle the problems of contemporary life.

REFERENCES AND FURTHER READINGS

Bangert-Drowns, R., & Bankert, E. (1990). *Meta-analysis of effects of explicit instruction for critical thinking.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Browne, M. N., & Keeley, S. M. (2004). *Asking the right questions: A guide to critical thinking* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Costa, A. (2001). *Developing minds: A resource book for teaching thinking*. Alexandria, VA: Association for Supervision and Curriculum Development.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53,* 449–455.

Halpern, D. F. (2001). *An on-line assessment of critical thinking skills.* Retrieved March 27, 2007, from http://berger.cmc.edu/rodinproject/

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking* (4th ed.). Mahwah, NJ: Erlbaum.

Hunt, E. B. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce*. New York: Russell Sage Foundation.

Jones, E. A., Hoffman, S., Moore, L. M., Ratcliff, G., Tibbetts, S., & Click, B. A. (1995). *National assessment of college student learning identifying college graduates' essential skills in writing, speech and listening, and critical thinking* (NCES 95-001). Washington, DC: U.S. Government Printing Office.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121–2234.

Marin-Burkhart, L. M., & Halpern, D. F. *Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains.* Unpublished manuscript.

Moseley, D., Baumfield, V., & Elliott, J. (2006). *Frameworks for thinking: A handbook for teaching and learning*. Cambridge, UK: Cambridge University Press.

National Center on Education and the Economy. (2007). *Tough choices or tough times: The report of the new commission on the skills of the American workforce*. Washington, DC: Author.

National Governor's Association. (2005). *Rate your future survey: A summary of the results from the National Governor's Association.* Retrieved from http://www.nga.org

Nisbett, R. (1993). *Rules for reasoning*. Hillsdale, NJ: Erlbaum.

Staib, S. (2003). Teaching and measuring critical thinking. *Journal of Nursing Education, 42,* 498–508.

Sternberg, R. J. (1997). *Successful intelligence*. New York: Plume.

Sternberg, R. J., Roediger, H. L., III, & Halpern, D. F. (Eds). (2007). *Critical thinking in psychology*. New York: Cambridge University Press.

52

ARTIFICIAL INTELLIGENCE

FRANCISCO ARCEDIANO
Auburn University

Artificial intelligence (AI) comprises a vast interdisciplinary field, which has benefited from its beginning from disciplines such as computer science, psychology, philosophy, neuroscience, mathematics, engineering, linguistics, economics, education, biology, control theory, and cybernetics. Although the goals of AI are as wide as the field is interdisciplinary, AI's main goal is the design and construction of automated systems (computer programs and machines) that perform tasks considered to require intelligent behavior (i.e., tasks that require adaptation to complex and changing situations).

The role of psychology in AI is twofold. On the one hand, psychologists can help in the development and construction of AI systems because knowledge of cognitive and reasoning processes such as perception, language acquisition, and social interaction is crucial to AI. AI has much to learn from humans because we are the best model of intelligent behavior we know, and because many AI machines will have to interact with us. On the other hand, psychologists could benefit from the AI techniques and tools to develop further their own discipline using AI tools such as modeling and simulation of theories, expert systems in diagnosis and organization, and interactive techniques in education, just to mention a few.

HISTORY OF AI

It seems that the desire to build machines that behave intelligently has always been a part of human history. For example, around 2500 BCE in Egypt, citizens and peregrines turned to oracles (statues with priests hidden inside) for advice. Homer's *Iliad*, a remarkable literature work from ancient Greece, narrates how the Greek god *Hephaestos* creates *Talos*, a man of bronze whose duty is to patrol and protect the beaches of Crete. The idea of building humans and machines with intelligence transferred from mythology into modern literature. For example, Karel Kapek's play *R.U.R.* (*Rossum's Universal Robots*), which opened in London in 1923, coined the word "robot." Shortly after, the very popular science fiction movie *Metropolis*, by Fritz Lang (1927), had a robot character (Maria) that played a decisive role in the plot of the movie. And, in the 1940s, Isaac Asimov started publishing his famous collection of books about robotics.

However, people not only wrote about the possibility of creating intelligent machines; they actually built them. For example, the ancient Greeks were fascinated with automata of all kinds, which they used mostly in theater productions and religious ceremonies for amusement. In the 4th century BCE, the Greek mathematician Archytas of Tarentum built a mechanical bird (a wooden pigeon) that, when propelled by a jet of steam or compressed air, could flap its wings. Supposedly, in one test, it flew a distance of 200 meters (however, once it fell to the ground, it could not take off again). Toward the end of the Middle Ages, clockmakers helped build devices that tried to mimic human and animal behavior. For example, Leonardo da Vinci built a humanoid automaton (an armored knight) around the end of the 15th century for the amusement of royalty. This armored knight was apparently able to make several humanlike motions, such as sitting up and moving its arms, legs, and neck. Reportedly, da Vinci also built a mechanical lion that could walk a programmable distance. In the early 16th century, Hans Bullmann created androids that could play musical instruments for the delight of paying customers. In the 18th century, Jacques de Vaucanson created a mechanical life-size figure (*The Flute Player*) capable of playing a flute with a repertoire of 12 different tunes. He also created an automatic duck (*The Digesting Duck*) that could drink, eat, paddle in water, and digest and excrete eaten grain.

In modern scientific AI, the first recognized work was Warren McCulloch and Walter Pitts's 1943 article *A Logical Calculus of the Ideas Immanent in Nervous Activity*, which laid the foundations for the development of neural networks. McCulloch and Pitt proposed a model of artificial neurons, suggesting that any computable function could be achieved by a network of connected neurons and that all logical connectives (and, or, not, etc.) could be implemented by simple network structures. In 1948, Wiener's popular book *Cybernetics* popularized the term *cybernetic* and defined the principle of the feedback theory. Wiener suggested that all intelligent behavior was the result of feedback mechanisms, or conditioned responses, and that it was possible to simulate these responses using a computer. One year later, Donald Hebb (1949) proposed a simple rule for modifying and updating the strength of the connections between neurons, which is now known as *Hebbian learning*. In 1950, Alan M. Turing published *Computing Machinery and Intelligence*, which was based on the idea that both machines and humans compute symbols and that this commonality should be the basis for building intelligent machines. Turing also introduced an operational strategy to test for intelligent behavior in machines based upon an imitation game known as the *Turing test*. (A brief description of the test and its impact on AI is discussed.) Because of the impact of his ideas on the field of AI, Turing is considered by many to be the father of AI.

The term AI was coined at the Dartmouth Summer Research Project on Artificial Intelligence in 1956 at Dartmouth College. This two-month workshop was organized by John McCarthy, Marvin Minsky, Claude Shannon, and Nathaniel Rochester and included as participants Trenchard More from Princeton, Arthur Samuel from IBM, Ray Solomonoff and Oliver Selfridge from MIT, and Allen Newell and Herbert Simon from Carnegie Tech, all of whom played fundamental roles in the development of AI. The Dartmouth workshop is considered the official birthplace of AI as a field, and it provided significant advances from previous work. For example, Allen Newell and Herbert Simon demonstrated a reasoning program, *the Logic Theorist*, which was capable of working with symbols and not just numbers.

The early years of AI were promising and full of successes. Both a symbolic approach (i.e., an approach that uses symbols and rules) and a subsymbolic approach (i.e., an approach that does not use rules but learns by itself) to AI coexisted with many successes. In the symbolic approach, some of the early successes include the presentation of the *General Problem Solver* by Newell, Shaw, and Simon (1963), a program designed to imitate human problem-solving protocols, and John McCarthy's *LISP* (1958), which became one of the predominant languages in AI. Some of the early successes in subsymbolic AI include the development of the *Adelines* by Widrow and Hoff (1960), which enhanced Hebbs's learning methods, and the *perceptron*, by Frank Rosenblatt (1962), which was the precursor of the artificial neural networks we know today.

However, by the end of the 1960s, difficulties arose as the AI promises from the decade before fell short and started to be considered "hype." Research in subsymbolic AI was mostly relegated after Minsky and Papert formally proved in 1969 that perceptrons (i.e., simple neural networks) were flawed in their representation mechanism because they could not represent the XOR (exclusive-OR) logical problem: a perceptron could not be trained to recognize situations in which either one or another set of inputs had to be present, but not both at the same time. The discovery that AI systems were not capable of solving simple logical problems that humans can easily solve resulted in significant reductions in research funding for artificial neural networks, and most researchers of this era decided to abandon the field.

The 1970s focused almost exclusively on different techniques in symbolic AI (such as production systems, semantic networks, and frames) and the application of these techniques to the development of expert systems (also known as knowledge-based systems), problem solving, and the understanding of natural language. By the mid-1980s, interest in symbolic AI began to decline because, once again, many promises remained unfulfilled. However, artificial neural networks became interesting again through what became known as the *connectionism movement*, largely due to two books discussing parallel distributed processing published by Rumelhart and McClelland (1986). These books demonstrated that complex networks could resolve the logical problems (e.g., X-OR) that early perceptrons could not resolve, and allowed networks to resolve many new problems. This new impulse of AI research resulted in the development of new approaches to AI during the late 1980s and 1990s, such as the subsymbolic approaches of evolutionary computing with evolutionary programming and genetic algorithms, behavior-based robotics, artificial life, and the development of the symbolic Bayesian networks. Today, AI is becoming successful in many different areas, especially in the areas of game playing, diagnosis, logistics planning, robotics, language understanding, problem solving, autonomous planning, scheduling, and control.

KNOWLEDGE REPRESENTATION

Knowledge representation addresses the problem of how knowledge about the world can be represented and what kinds of reasoning can be done with that knowledge. Knowledge representation is arguably the most relevant topic in AI because what artificial systems can do depends on their ability to represent and manipulate knowledge. Traditionally, the study of knowledge representation has had two different approaches: symbolic AI (also known as the top-down approach) and subsymbolic AI (also known as the bottom-up approach).

Symbolic AI (Top-Down Approach)

Symbolic AI has been referred to also as conventional AI, classical AI, logical AI, neat AI, and Good Old Fashioned AI (GOFAI). The basic assumption behind symbolic AI is that (human) knowledge can be represented explicitly in a declarative form by using facts and rules: Knowledge, either declarative (or explicit) or procedural (or implicit), can be described by using symbols and rules for their manipulation.

Symbolic AI is traditionally associated with a top-down approach because it starts with all the relevant knowledge already present for the program to use and the set of rules to decompose the problem through some inferential mechanism until reaching its goal. A top-down approach can be used only when we know how to formalize and operationalize the knowledge we need to solve the problem. Because of its higher level of representation, it is well suited to perform relatively high-level tasks such as problem solving and language processing. However, this approach is inherently poor at solving problems that involve ill-defined knowledge, and when the interactions are highly complex and weakly interrelated, such as in commonsense knowledge, when we do not know how to represent the knowledge hierarchy, or when we do not know how to represent the mechanism needed to reach a solution. Many different methods have been used in symbolic AI.

Predicate Logic or First-Order Logic

Logic is used to describe representations of our knowledge of the *world*. It is a well-understood formal language, with a well-defined, precise, mathematical syntax, semantics, and rules of inference. Predicate logic allows us to represent fairly complex facts about the world, and to derive new facts in a way that guarantees that, if the initial facts are true, so then are the conclusions. There are well-defined procedures to prove the truth of the relationships and to make inferences (substitution, modus ponens, modus tollens, unification, among others).

Rule-Based Systems (Production Systems)

A rule-based system consists of a set of IF-THEN rules, a set of facts, and some interpreter controlling the application of the rules, given the facts. A rule-based system represents knowledge in terms of a set of rules that guides the system inferences given certain facts (e.g., IF the temperature is below 65 F degrees AND time of day is between 5:00 p.m. and 11:00 p.m. THEN turn on the heater). Rule-based systems are often used to develop *expert systems*. An expert system contains knowledge derived from an expert or experts in some domain, and it exhibits, within that specific domain, a degree of expertise in problem solving that is comparable to that of the human experts. Simply put, an expert system contains a set of IF-THEN rules derived from the knowledge of human experts. Expert systems are supposed to support inspection of their reasoning processes, both in presenting intermediate steps and in answering questions about the solution processes: At any time we can inquire why expert systems are asking certain questions, and these systems can explain their reasoning or suggested decisions.

Fuzzy Logic

Fuzzy logic is a superset of classical Boolean logic that has been extended to handle the concept of partial truth. One of the limitations of predicate (first-order) logic is that it relies in Boolean logic, in which statements are entirely true or false. However, in the real world, there are many situations in which events are not clearly stated and the truth of a statement is a matter of degree (e.g., if someone states a person is tall, the person can be taller than some people but shorter than other people; thus, the statement is true only sometimes). Fuzzy logic is a continuous form of logic that uses modifiers to describe different levels of truth. It was specifically designed to represent uncertainty and vagueness mathematically and provide formalized tools for dealing with the imprecision intrinsic to many problems. Because fuzzy logic can handle approximate information systematically, it is ideal for controlling and modeling complex systems in which an inexact model exists or systems where ambiguity or vagueness is common. Today, fuzzy logic is found in a variety of control applications such as expert systems, washing machines, video cameras (e.g., focus aperture), automobiles (e.g., operation of the antilock braking systems), refrigerators, robots, failure diagnosis, pattern classifying, traffic lights, smart weapons, and trains.

Semantic Networks, Frames, and Scripts

Semantic networks are graphical representations of information consisting of nodes, which represent an object or a class, and links connecting those nodes, representing the attributes and relations between the nodes. Semantic networks are often called *conceptual graphs*. Researchers originally used them to represent the meaning of words in programs that dealt with natural language processing (e.g., understanding news), but they have also been applied to other areas, such as modeling memory.

One interesting feature of semantic networks is how convenient they are to establish relations between different areas of knowledge and to perform inheritance reasoning. For example, if the system knows that entity A is human, then it knows that all human attributes can be part of the description of A (inheritance reasoning).

A problem with semantic networks is that as the knowledge to be represented becomes more complex, the representation grows in size, it needs to be more structured, and it becomes hard to define by graphical representations. To allow more complex and structured knowledge representation, frames were developed. A frame is a collection of attributes or slots with associated values that describe some real-world entity. Frame systems are a powerful way of encoding information to support reasoning. Each frame represents a class or an instance (an element of a class, such as height) and its slots represent an attribute with a value (e.g., seven feet).

Scripts are used to develop ideas or processes that represent recurring actions and events. They are often built on semantic networks or frames, although production systems are also common. Scripts are used to make inferences on a whole set of actions that fall into a stereotypical pattern. A script is essentially a prepackaged inference chain relating to a specific routine situation. They capture knowledge about a sequence of events, and this knowledge has been used as a way of analyzing and describing stories. Typical examples of scripts are the sequence of actions and the knowledge needed to, for example, take a flight or buy a train ticket.

Bayesian Networks

Bayesian networks are also known as belief networks, probabilistic networks, causal networks, knowledge maps, or graphical probability models. They are a probabilistic graphical model with nodes, which represent discrete or continuous variables, and links between those nodes, which represent the conditional dependencies between variables. This graphical representation with nodes and links connecting the nodes provides an intuitive graphical visualization of the knowledge, including the interactions among the various sources of uncertainty. Because a Bayesian network is a complete model for the variables and their relationships, it can be used to answer probabilistic queries about them, and it can allow us to model and reason about uncertainty in complex situations. For example, a Bayesian network can be used to calculate the probability of a patient having a specific disease, given the absence or presence of certain symptoms, if the probabilistic dependencies between symptoms and disease are assumed to be known.

Bayesian networks have been used for diverse applications, such as diagnosis, expert systems, planning, learning, decision making, modeling knowledge in bioinformatics (gene regulatory networks, protein structure), medicine, engineering, document classification, image processing, data fusion, decision support systems, and e-mail spam filtering.

Subsymbolic AI (Bottom-Up Approach)

After researchers became disillusioned in the mid-1980s with the symbolic attempts at modeling intelligence, they looked into other possibilities. Some prominent techniques arose as alternatives to symbolic AI, such as connectionism (neural networking and parallel distributed processing), evolutionary computing, particle swarm optimization (PSO), and behavior-based AI.

In contrast with symbolic AI, subsymbolic AI is characterized by a bottom-up approach to AI. In this approach, the problem is addressed by starting with a relatively simple abstract program that is built to learn by itself, and it builds knowledge until reaching an optimal solution. Thus, it starts with simpler elements and then, by interacting with the problem, moves upwards in complexity by finding ways to interconnect and organize the information to produce a more organized and meaningful representation of the problem.

The bottom-up approach has the advantage of being able to model lower-level human and animal functions, such as vision, motor control, and learning. It is more useful when we do not know how to formalize knowledge and we do not know how to reach the answer beforehand.

Neural Networks

Neural networks, or more correctly, artificial neural networks, to differentiate them from biological neural networks, are computing paradigms loosely modeled after the neurons in the brain and designed to model or mimic some properties of biological neural networks. They consist of interconnected processing elements called *nodes* or *neurons* that work together to produce an output function. The output of a neural network depends on the cooperation of the individual neurons within the network. Because it relies on its collection of neurons to perform or reach a solution, it can still perform its overall function even if some of the neurons are not functioning, which makes them able to tolerate error or failure. They learn by adapting their structure based on external or internal information that flows through the network. Thus, they are mostly used to model complex relationships between inputs and outputs or to find patterns in data.

As you may recall, one of the simplest instantiations of a neural network, the perceptrons, were very popular in the early 1960s (Rosenblatt, 1962), but interest in them dwindled at the end of the 1960s because they were not able to represent some simple logical problems (Minsky & Papert, 1969). They became hugely popular again in the mid-1980s (McClelland et al., 1986; Rumelhart et al., 1986) because the problem associated with the perceptron was addressed and because of the decreased interest in symbolic approaches to AI. The utility of artificial neural network models lies in the fact that they can be used to infer a function from observations. Such inferences are particularly useful in applications in which the complexity of the data or task makes the design of such a function by hand impractical.

There are three major learning paradigms in the neural network field, each corresponding to a particular abstract learning task: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the network is given a set of examples (inputs) with the correct responses (outputs), and it finds a function that matches these examples with the responses. The network infers the mapping implied by the data by using the mismatch between the mapping and the data and correcting the weight of the connection between the nodes until the network is able to match the inputs with the outputs. With this function, new sets of data or stimuli, previously unknown to the system, can be correctly classified. In unsupervised learning, the network has to learn patterns or regularities in the inputs when no specific output values are supplied or taught to the network. Finally, in reinforcement learning, the set of examples with their answers are not given to the network, but are found by interacting with the environment and integrating the instances that lead to reinforcement. The system performs an action and, based on the consequences of that action (some cost according to some dynamics), relates the stimuli (inputs) and the responses (outputs). Artificial neural networks have been applied successfully to speech recognition, image analysis, adaptive control and navigation, video games, autonomous robots, decision making, detection of credit card fraud, data mining, and pattern recognition, among other areas.

Evolutionary Computation

Evolutionary computation applies biologically inspired concepts from natural selection (such as populations, mutation, reproduction, and survival of the fittest) to generate increasingly better solutions to the problem. Some of the most popular methods are evolutionary programming, evolution strategies, and genetic algorithms. Evolutionary computation has been successfully applied to a wide range of problems including aircraft design, routing in communications networks, game playing, robotics, air traffic control, machine learning, pattern recognition, market forecasting, and data mining. Although evolutionary programming, evolution strategies, and genetic algorithms are similar at the highest level, each of these varieties implements evolutionary algorithms in a different manner.

Evolutionary programming, originally conceived by Lawrence J. Fogel in 1960, emphasizes the relationship between parent solutions (the solutions being analyzed) and their offspring (new solutions resulting from some modification of the parent solutions). Fogel, Owens, and Walsh's 1966 book *Artificial Intelligence Through Simulated Evolution* is the landmark publication in this area of AI. In general, in evolutionary programming, the problem to be solved is represented or encoded in a string of variables that defines all the potential solutions to the problem. Each full set of variables with its specific values is known as an *individual* or *candidate solution*. To solve the problem, a population of "individuals" is created, with each individual representing a random possible solution to the problem. Each of the individuals (i.e., each candidate solution) is evaluated and assigned a fitness value based on how effective the candidate solution is to solving the problem. Based on this fitness value, some individuals (usually the most successful) are selected to be parents, and offspring are generated from these parents.

In the generation process, a mutation operator selects elements of the parents' representation of the solution and manipulates these elements when they are transferred to the offspring. A mutation operator is a rule that selects random variables and randomly alters the value of these variables in some degree, generating new individuals or candidate solutions from the selected parents. Thus, some characteristics of the parent solutions are changed slightly and then transferred to the offspring solution. In general, the degree of mutation is greater in the first generations, and it is gradually decreased as generations evolve and get closer to an optimal solution. The offspring candidate solutions are then evaluated based on their fitness, just like their parents were, and the process of generating offspring from the parents is repeated until an individual with sufficient quality (an optimal solution to the problem) is found or a previously determined computational limit is reached (e.g., after evolving for a given number of generations).

Evolution strategies (Bäck, Hoffmeister, & Schwefel, 1991; Rechenberg, 1973) and evolutionary programming share many similarities. The main difference is that, in evolution strategies, offspring are generated from the selected parents not only by using a mutation operator but also by recombination of the code from selected parents through a crossover operator. A crossover operator applies some rule to recombine the elements of the selected parents to generate offspring. The recombination operation simulates some reproduction mechanism to transfer elements from the parents to their offspring.

The crossover operator can take many variants (e.g., interchange the first half of the elements from one of the parents and the second half from the other one for one offspring; the reverse for another offspring). The crossover operator is inspired by the role of sexual reproduction in the evolution of living things. The mutation operator is inspired by the role of mutation in natural evolution. Generally, both mutation and reproduction are used simultaneously. Recombination and mutation create the necessary diversity and thereby facilitate novelty, while selection acts as a force increasing quality.

Genetic algorithms, popularized by John Holland (1975), are similar to evolution strategies in the general steps that the algorithm follows. However, there are substantial differences in how the problem is represented. One of the main differences is that the problem to be solved is encoded in each individual of the population by having arrays of bits (bit-strings), which represent chromosomes. Each bit in the bit-string is analogous to a gene (i.e., an element that represents a variable or part of a variable of the problem). In a genetic algorithm, each individual or candidate solution is encoded at a genotype level, whereas in evolutionary programming and evolution strategy, the problem is encoded at a phenotype level, in which there is a one-to-one relationship between each value encoded in the phenotype and the real value that it represents in the problem. A genetic algorithm can differentiate between genotype (the genes) and phenotype (the expression of a collection of genes). The manipulation at the level of genotype allows for more elaborated implementation of the crossover operator and the mutation operator. Additionally, the focus on genetic algorithms when creating offspring in successive generations is on reproduction (crossover) and no mutation, which is often considered as a background operator or secondary process.

Particle Swarm Optimization

PSO applies the concept of social interaction to problem solving. It was developed in 1995 by James Kennedy and Russ Eberhart, and it is inspired by the social behavior of bird flocking and fish schooling.

PSO shares many similarities with evolutionary computation techniques. The system is initialized with a population of random potential solutions (known in this framework as particles), which searches for an optimal solution to the problem by updating generations. However, unlike evolutionary computing, PSO has no evolution operators such as crossover and mutation. In PSO, the particles *fly* through the problem space by following the current best solution (the particle with the best fitness). Each particle (individual) records its current position (location) in the search space, records the location of the best solution found so far by the particle, and a gradient (direction) in which the particle will travel if undisturbed. In order to decide whether to change direction and in which direction to travel (searching for the optimal solution), the particles work with two fitness values: one for that specific particle, and another for the particle closer to the solution (best candidate solution). Thus, particles can be seen as simple agents that fly through the search space and record (and possibly communicate) the best solution that they have discovered so far. Particles travel in the search space by simply adding their vector (the direction in which they were traveling) and the vector (direction) of the best solution candidate. Then, each particle computes its new fitness, and the process continues until the particles converge on an optimal solution.

Behavior-Based AI

Behavior-based AI is a methodology for developing AI based on a modular decomposition of intelligence. It was made famous by Rodney Brooks at MIT (see Brooks, 1999, for a compendium of his most relevant papers in the topic), and it is a popular approach to building simple robots, which surprisingly appear to exhibit complex behavior. The complexity of their behavior lies in the perception of the observer, not in the processing mechanism of the system. This approach questions the need for modeling intelligence using complex levels of knowledge representation and the need for a higher cognitive control. Brooks presents a series of simple robots that mimic intelligent behavior by using a set of independent semi-autonomous modules, which interact independently with the environment and are not communicating information at any higher level. For example, a spiderlike robot will navigate through a path with obstacles just by each of its legs addressing its own situation, without a mechanism relating what each other leg *knows* about the environment. This approach has been successful when dealing with dynamic, unpredictable environments. Although behavior-based AI has been popular in robotics, it also can be applied to more traditional AI areas.

AI IN PSYCHOLOGY

The greatest impact of AI in psychology has been through the development of what has come to be known as the *information processing paradigm* or the *computer analogy*. Once computers started to be perceived as information-processing systems able to process symbols and not just numbers, an analogy between computers and the human mind was established. Both systems receive inputs (either through sensors or as the output from other processes or devices), process those inputs through a central processing unit, and generate an output (through motor responses or as the input for other processes or devices). The idea that the mind works on the brain just as a program works on a computer is the focus of cognitive psychology, which is concerned with information-processing mechanisms focusing especially on processes such as attention, perception, learning, and memory. It is also concerned with the structures and representations involved in cognition in general.

One of the dominant paradigms in psychology, before the surge of AI at the end of the 1950s, was behaviorism, in which the focus was on the study of the responses of the organism (behavior) given particular inputs (stimuli). Its main assumption was that, because researchers can only scientifically study what they can observe and measure, behavior should be the only subject matter of study of scientific psychology. With cognitive psychology and the computer analogy, the focus started to shift toward the study of mental processes, or *cognition*. Cognitive psychology is interested in identifying in detail what happens between stimuli and responses. To achieve this goal, psychological experiments need to be interpretable within a theoretical framework that describes and explains mental representations and procedures. One of the best ways of developing these theoretical frameworks is by forming and testing computational models intended to be analogous to mental operations. Thus, cognitive psychology views the brain as an information-processing device that can be studied through experimentation and whose theories can be rigorously tested and discussed as computer programs.

A stronger focus on computer modeling and simulations and on study of cognition as a system resulted in the development of *cognitive science*. Cognitive science is an interdisciplinary discipline concerned with learning how humans, animals, and machines acquire knowledge, represent that knowledge, and how those representations are manipulated. It embraces psychology, artificial intelligence, neuroscience, philosophy, linguistics, anthropology, biology, evolution, and education, among other disciplines.

More recently, AI philosophy and techniques have impacted a new discipline, *cognitive neuroscience*, which attempts to develop mathematical and computational theories and models of the structures and processes of the brain in humans and other animals. This discipline is concerned directly with the nature of the brain and tries to be more biologically accurate by modeling the behavior of neurons, simulating, among other things, the interactions among different areas of the brain and the functioning of chemical pathways. Cognitive neuroscience attempts to derive cognitive-level theories from different types of information, such as computational properties of neural circuits, patterns of behavioral damage as a result of brain injury, and measures of brain activity during the performance of cognitive tasks.

A new and interesting approach resulting from developments in computer modeling is the attempt to search and test for a so-called *unified architecture* (a unified theory of cognition). The three most dominant unified theories of cognition in psychology are SOAR (based on symbolic AI), PDP (based on subsymbolic AI), and ACT-R (originally based on symbolic AI, currently a hybrid, using both symbolic and subsymbolic approaches).

SOAR (State Operator and Result)

SOAR (Laird, Newell, & Rosenbloom, 1987; Newell, 1990) describes a general cognitive architecture for developing systems that exhibit intelligent behavior. It represents and uses appropriate forms of knowledge such as procedural, declarative, and episodic knowledge. It employs a full range of problem-solving methods, and it is able to interact with the outside world.

PDP (Parallel Distributed Processing)

In the 1980s and 1990s, James L. McClelland, David E. Rumelhart, and the PDP Research Group (McClelland et al., 1986; Rumelhart et al., 1986) popularized artificial neural networks and the connectionist movement, which had lain dormant since the late 1960s. In the connectionist approach, cognitive functions and behavior are perceived as emergent processes from parallel, distributed processing activity of interconnected neural populations, with learning occurring through the adaptation of connections among the participating neurons. PDP attempts to be a general architecture and explain the mechanisms of perception, memory, language, and thought.

ACT-R (Adaptive Control of Thought-Rational)

In its last instantiation, ACT-R (Anderson et al., 2004; Anderson & Lebiere, 1998) is presented as a hybrid cognitive architecture. Its symbolic structure is a production system. The subsymbolic structure is represented by a set of massive parallel processes that can be summarized by a number of mathematical equations. Both representations work together to explain how people organize knowledge and produce intelligent behavior. ACT-R theory tries to evolve toward a system that can perform the full range of human cognitive tasks, capturing in great detail how we perceive, think about, and act on the world. Because of its general architecture, the theory is applicable to a wide variety of research disciplines, including perception and attention, learning and memory, problem solving and decision making, and language processing.

The benefits of applying computer modeling to psychological hypotheses and theories are multiple. For example, computer programs provide unambiguous formulations of a theory as well as means for testing the sufficiency and consistency of its interconnected elements. Because computer modeling involves using a well-formulated language, it eliminates vagueness and highlights hidden or ambiguous intermediated processes that were not previously known or made explicit with the verbal description of the theory. Explicit formulations also allow researchers to falsify (i.e., test) the theory's assumptions and conclusions. Additionally, given the same data set, alternative programs/theories can be run through the data to analyze which hypotheses are more consistent with the data and why.

Computer modeling has focused mostly in the areas of perception, learning, memory, and decision making, but it is also being applied to the modeling of mental disorders (e.g., neurosis, eating disorders, and autistic behavior), cognitive and social neuroscience, scientific discovery, creativity, linguistic processes (e.g., dyslexia, speech movements), attention, and risk assessment, among other fields. Importantly, the impact of AI on psychology has not been relegated to only theoretical analyses. AI systems can be found in psychological applications such as the development of expert systems for clinical diagnosis and education, human-computer interaction and user interfaces, and many other areas, although the impact in application areas is not as extended as might be expected—there seems to be much work to do in this area.

CRITICISMS OF AI

Is true AI possible? Can an AI system display true intelligence and consciousness? Are these functions reserved only to living organisms? Can we truly model knowledge?

Weak AI Versus Strong AI

The weak AI view (also known as *soft-cautious AI*) supports the idea that machines can be programmed to act as if they were intelligent: Machines are only capable of simulating intelligent behavior and consciousness (or understanding), but they are not really capable of true understanding. In this view, the traditional focus has been on developing machines able to perform in a specific task with no intention of building a complete system able to perform intelligently in all or most situations.

Strong AI (also known as *hard AI*) supports the view that machines are really intelligent, and that, someday, they could have understanding and conscious minds. This view assumes that all mental activities of humans can be eventually reducible to algorithms and processes that can be implemented in a machine. Thus, for example, there should be no fundamental differences between a machine that emulates all the processes in the brain and the actions of a human being, including understanding and consciousness.

One of the problems with the strong AI view centers on the following two questions: (a) How do we know that an artificial system is truly intelligent? (b) What makes a system (natural or not) intelligent? Even today, there is no clear consensus on what *intelligence* really is. Turing (1950) was aware of this problem and, recognizing the difficulties in agreeing on a common definition on intelligence, proposed an operational test to circumvent this question. He named this test *the imitation game* and it was later known as the *Turing test*.

The Turing test is conducted with two people and a machine. One person, who sits in a separate room from the machine and the other person, plays the role of an interrogator or judge. The interrogator knows the person and the machine only as A and B, respectively, and has no way of knowing beforehand which is the person and which is the machine. The interrogator can ask A and B any question she wishes. The interrogator's aim is to determine which one is the person and which one is the machine. If the machine fools the interrogator into thinking that it is a person, then we can conclude that the machine can think and it is truly intelligent, or at least as intelligent as the human counterpart. The Turing test has become relevant in the history of AI because some of the criticisms and philosophical debates about the possibilities of AI have focused on this test.

Criticisms of the Strong AI View

For many decades it has been claimed that Gödel's (1931) *incompleteness theorem* precludes the possibility of the development of a true artificial intelligence. The idea behind the incompleteness theorem is that within any given branch of mathematics, there will always be some propositions that cannot be proved or disproved within the system using the rules and axioms of that mathematical branch itself. Thus, there will be principles that cannot be proved in a system, and the system cannot deduce them. Because machines do little more than follow a set of rules, they cannot truly emulate human behavior. Human behavior is too complex to be described by any set of rules.

Machines Cannot Have Understanding—The Chinese Room Argument

John Searle stated in his *Chinese room argument* that machines work with encoded data that describe other things, and that data are meaningless without a cross-reference to the things described. This point led Searle to assert that there is no meaning or understanding in a computational machine itself. As a result, Searle claimed that even a machine that passes the Turing test would not necessarily have understanding or be conscious. Consciousness seems necessary in order to show understanding.

In his *thought experiment,* Searle asked people to imagine a scenario in which he is locked in a room and receives some Chinese writing (e.g., a question). He does not know Chinese, but he is given a rule book with English instructions that allows him to correlate the set of Chinese symbols he receives with another set of Chinese symbols (e.g., by their shape), which he would give back as an answer. Imagine the set of rules to correlate both sets of symbols is so advanced that it allows him to give a meaningful answer to the question in Chinese. Imagine the same thing happening with an English question and its answer. For a Chinese and an English observer, respectively, the answers from Searle are equally satisfying and meaningful (i.e., intelligent). In both situations Searle would pass the Turing's test. However, can we say Searle understands Chinese in the same way that he understands English? Actually, he does not know any Chinese. A computer program would behave similarly by taking a set of formal symbols (Chinese characters) as input and, following a set of rules in its programming, correlating them with another set of formal symbols (Chinese characters), which it presents as output. In this thought experiment, the computer would be able to pass the Turing test. It converses with Chinese speakers, but they do not realize they are talking with a machine. From a human observer's view, it seems that the computer truly understands Chinese. However, in the same way that Searle does not understand Chinese, the machine does not understand Chinese either. What the computer does is mindless manipulation of the symbols, just as Searle was doing. Although it would pass the Turing test, there is no genuine understanding involved.

Several other authors have argued against the possibility of a strong AI (see, e.g., Penrose, 1989). The arguments denying the possibility of a strong AI and their counterarguments have populated the AI literature, especially the philosophical discussion of AI.

Machines able to sense and acquire experiences by interacting with the world should be able to solve most of these criticisms. In fact, most AI researchers are concerned with specific issues and goals to resolve, such as how to improve actual systems or how to create new approaches to the problem of AI, and not passing the Turing test and proving that AI systems can be truly intelligent. Additionally, the Turing test has been seriously criticized as a valid test for machine intelligence. For example, Ford and Hayes (1998) highlighted some of the problems with the Turing test. The central defect of the Turing test is its species-centeredness. It assumes that human thought is the highest achievement of thinking against which all other forms of thinking must be judged. The Turing test depends on the subjectivity of the judge, and it is culture-bound (a conversation that passes the test in the eyes of a British judge might fail it according to a Japanese or Mexican judge). More important, it does not admit as measures of intelligence weaker, different, or even stronger forms of intelligence than those deemed human. Ford and Hayes made this point clearly when they compared AI with artificial flight. In the early thinking about flight, success was defined as the imitation of a natural model: for flight, a bird; for intelligence, a human. Only when pioneers got some distance from the model of a bird did flying become successful. In the same analogy, you do not deny an airplane can fly just because it does not fly like a bird does. However, truly artificial intelligence is denied if it does not parallel that of a human being.

Many supporters of strong AI consider that it is not necessary to try to isolate and recreate consciousness and understanding specifically. They accept that consciousness could be a by-product of any sufficiently complex intelligent system, and it will emerge automatically from complexity. They focus on the analogy of the human brain. In humans, a single neuron has nothing resembling intelligence. Only when billions of neurons combine to form a mind does intelligence emerge. It would appear, then, that the brain is more than the sum of its parts. Intelligence emerges with sufficient joint complexity of neurons.

In the final analysis, the debate of the possibility of a strong AI seems to be based on the classical mind/body problem. On the one side, we have the physicalism view (the belief that nothing exists but the physical world that can be studied scientifically; mental states are assumed to mirror brain states). In this view, strong AI could be eventually possible. On the other side, we have the view based on dualism (mind and matter are not the same thing). In this view, strong AI is not possible.

SUMMARY

The most important influence of AI in psychology has been the computer metaphor, in which both living organisms and computers can be understood as information processors. The information-processing approach has brought psychology new paradigms such as cognitive psychology and cognitive science. The goal of these disciplines is to learn what happens within the organism, at the level of the brain. Brain or mental processes cannot be observed directly, but computer formalization of theories of how mental processes work can be tested scientifically.

Although AI and cognitive science share techniques and goals, there is at least one fundamental difference between them. Psychology has the restriction that its computer programs and simulations have to achieve the same results as the simulated system (human and animal cognition), and they have to do it following the same processes. Only when there is a match with real processes can psychologists assume that the relationships proposed in the program were correct. AI does not have a restriction of similarity of processes: Psychological theory must be able to predict error empirically. AI focuses on efficiency; psychology focuses on plausibility.

Additionally, computer modeling and the focus on brain structure and their function have given rise to the field of neuroscience. Neuroscience is one of the most rapidly emerging disciplines in psychology, largely due to computer modeling. Computer modeling allows for the formalization of models and the testing of hypotheses about how neurons work together. The new computational techniques applied to brain activity measurement (e.g., fMRI) allow researchers to observe how brain structures actually work. Thus, formal models that represent theories and hypotheses can be tested against real data.

The impact of cognitive modeling in psychology should increase as psychological models move forward toward a unified theory of cognition. Only through computer formalizations and implementations does it seem plausible to work on a unified view of psychology in which all the theories might be integrated in a single framework. One of the underlying assumptions of computer modeling is that the role of formal modeling of theories and hypotheses in psychology plays a role similar to the one of mathematics in the physical sciences.

Although computer modeling has strongly influenced AI in psychology, the impact of AI in psychology goes far beyond computer modeling. AI techniques can be applied to many areas of psychology, especially those that focus on diagnosis, classification, and decision making.

REFERENCES AND FURTHER READINGS

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*, 1036–1060.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Bäck, T., Hoffmeister, F., & Schwefel, H.-P. (1991). A survey of evolution strategies, In R. K. Belew & L. B. Booker (Eds.), *Proceedings of the fourth international conference on genetic algorithms* (pp. 2–9). San Mateo, CA: Morgan Kaufmann.

Boden, M. A. (Ed.). (1988). *Computer models of mind.* Cambridge, UK: Cambridge University Press.

Boden, M. A. (Ed.). (1989). *Artificial intelligence in psychology: Interdisciplinary essays*. Cambridge, MA: MIT Press.

Boden, M. A. (Ed.). (1996). *Artificial intelligence*. San Diego, CA: Academic Press.

Brooks, R. A. (1999). *Cambrian intelligence*. Cambridge, MA: MIT Press.

Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial intelligence through simulated evolution*. New York: Wiley.

Ford, K. M., & Hayes, P. J. (1998). On computational wings: Rethinking the goals of artificial intelligence. *Scientific American, 9,* 78–83.

Gödel, K. (1931). Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik, 38*, 173–198

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks* (Vol. 4, pp. 1942–1948). Piscataway, NJ: IEEE Press.

Laird, J. E., Newell, A., & Rosembloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence, 33*, 1–64.

McCarthy, J. (1958). Programs with common sense. In *Proceedings of the Symposium on Mechanization of Thought Processes* (Vol. 1, pp. 77–84). London: Her Majesty's Stationery Office.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115–137.

Minsky, M. L., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.

Newell, A. (1990). *Unified theories of cognition. The William James lectures*. Cambridge, MA: Harvard University Press.

Newell, A., Shaw, J. C., & Simon, H. (1963). GPS, A program that simulates human thought. In E. A. Feingenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 279–296). New York: McGraw-Hill. (Reprinted from *Lernende Automaten*, pp. 109–124, by H. Billing, Ed., Munich: Oldenbourg)

Penrose, R. (1989). *The emperor's new mind*. Oxford, UK: Oxford University Press.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution*. Stuttgart, Germany: Fromman-Holzboog.

Rich, E., & Knight, K. (1991). *Artificial intelligence* (2nd ed.). New York: McGraw-Hill.

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanism.* Chicago: Spartan.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: A modern approach* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.

Wagman, M. (1993). *Cognitive psychology and artificial intelligence. Theory and research in cognitive science*. Westport, CT: Praeger.

Wagman, M. (1995). *The sciences of cognition: Theory and research in psychology and artificial intelligence*. Westport, CT: Praeger.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON Convention Record* (Part 4, pp. 96–104). New York: IRE.

Wiener, N. (1948). *Cybernetics.* New York: Wiley.