

Epistemology

An Anthology

Edited by

Ernest Sosa and Jaegwon Kim

Brown University

With the assistance of

Matthew McGrath

Texas A&M University



Blackwell
Publishing

BLACKWELL PHILOSOPHY ANTHOLOGIES

Each volume in this outstanding series provides an authoritative and comprehensive collection of the essential primary readings from philosophy's main fields of study. Designed to complement the *Blackwell Companions to Philosophy* series, each volume represents an unparalleled resource in its own right, and will provide the ideal platform for course use.

- 1 Cottingham: *Western Philosophy: An Anthology*
- 2 Cahoon: *From Modernism to Postmodernism: An Anthology*
- 3 LaFollette: *Ethics in Practice: An Anthology*
- 4 Goodin and Pettit: *Contemporary Political Philosophy: An Anthology*
- 5 Eze: *African Philosophy: An Anthology*
- 6 McNeill and Feldman: *Continental Philosophy: An Anthology*
- 7 Kim and Sosa: *Metaphysics: An Anthology*
- 8 Lycan: *Mind and Cognition: An Anthology* (second edition)
- 9 Kuhse and Singer: *Bioethics: An Anthology*
- 10 Cummins and Cummins: *Minds, Brains, and Computers – The Foundations of Cognitive Science: An Anthology*
- 11 Sosa and Kim: *Epistemology: An Anthology*

© 2000 by Blackwell Publishing Ltd

350 Main Street, Malden, MA 02148-5018, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton South, Melbourne, Victoria 3053, Australia
Kurfürstendamm 57, 10707 Berlin, Germany

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs, and Patents Act 1988, without the prior permission of the publisher.

First published 2000 by Blackwell Publishing Ltd
Reprinted 2001, 2002, 2003

Library of Congress Cataloging-in-Publication Data

Epistemology: an anthology / edited by Ernest Sosa and Jaegwon Kim;
with the assistance of Matthew McGrath.
p. cm.—(Blackwell philosophy anthologies; 11)
Includes bibliographical references and index.
ISBN 0-631-19723-0 (alk. paper)—ISBN 0-631-19724-9 (pbk. : alk. paper)
1. Knowledge, Theory of. I. Sosa, Ernest. II. Kim, Jaegwon.
III. McGrath, Matthew. IV. Series.
BD161.E615 2000
121—dc21

99-16132
CIP

A catalogue record for this title is available from the British Library.

Set in 9 on 11pt Ehrhardt
by Kolam Information Services Pvt Ltd., Pondicherry, India
Printed and bound in the United Kingdom
by MPG Books Ltd., Bodmin, Cornwall

For further information on
Blackwell Publishing, visit our website:
<http://www.blackwellpublishing.com>

Contents

Preface	ix
Acknowledgments	x
Part I Skepticism	1
Introduction	3
1 The Problem of the External World Barry Stroud	6
2 Proof of an External World G. E. Moore	24
3 Four Forms of Scepticism G. E. Moore	27
4 Certainty G. E. Moore	29
5 Skepticism, Naturalism and Transcendental Arguments P. F. Strawson	33
6 An Argument for Skepticism Peter Unger	42
Part II Defining Knowledge	53
Introduction	55
7 Is Justified True Belief Knowledge? Edmund Gettier	58
8 A Proposed Definition of Propositional Knowledge Peter Klein	60

Contents

9	Selections from <i>Thought</i> Gilbert Harman	67
10	Knowledge and Skepticism Robert Nozick	79
Part III Contemporary Foundationalism and Coherentism		103
	Introduction	105
11	The Myth of the Given Roderick M. Chisholm	107
12	Does Empirical Knowledge Have a Foundation? Wilfrid Sellars	120
13	Epistemic Principles Wilfrid Sellars	125
14	The Raft and the Pyramid Ernest Sosa	134
15	A Coherence Theory of Truth and Knowledge Donald Davidson	154
Part IV Epistemic Justification		165
	Introduction	167
16	Evidentialism Richard Feldman and Earl Conee	170
17	Skepticism and Rationality Richard Foley	182
18	Epistemic Norms John Pollock	192
19	A Foundherentist Theory of Empirical Justification Susan Haack	226
Part V The Pyrrhonian Problematic		237
	Introduction	239
20	Foundationalism, Epistemic Principles, and the Cartesian Circle James Van Cleve	242
21	Can Empirical Knowledge Have a Foundation? Laurence Bonjour	261
22	Reflective Knowledge in the Best Circles Ernest Sosa	274
Part VI Epistemology Naturalized		287
	Introduction	289

	Contents	
23	Epistemology Naturalized W. V. Quine	292
24	What Is “Naturalized Epistemology”? Jaegwon Kim	301
25	Why Reason Can’t Be Naturalized Hilary Putnam	314
26	The Old Skepticism, the New Foundationalism, and Naturalized Epistemology Robert Audi	325
 Part VII Epistemic Externalism		335
	Introduction	337
27	What Is Justified Belief? Alvin I. Goldman	340
28	How to Think about Reliability William P. Alston	354
29	The Generality Problem for Reliabilism Earl Conee and Richard Feldman	372
30	Externalism and Epistemology Naturalized Keith Lehrer	387
31	Externalism and Skepticism Richard Fumerton	401
32	Knowledge and the Internal John McDowell	413
33	Knowledge and the Social Articulation of the Space of Reasons Robert Brandom	424
 Part VIII Virtue Epistemology and Proper Cognitive Functioning		433
	Introduction	435
34	Epistemic Folkways and Scientific Epistemology Alvin I. Goldman	438
35	Warrant: A First Approximation Alvin Plantinga	445
36	Virtues of the Mind Linda Zagzebski	457
37	Virtues and Vices of Virtue Epistemology John Greco	468
 Part IX Epistemic Contextualism		477
	Introduction	479
38	Solving the Skeptical Problem Keith DeRose	482

Contents

39	Elusive Knowledge David Lewis	503
40	Contextualist Solutions to Epistemological Problems: Scepticism, Gettier, and the Lottery Stewart Cohen	517
Part X Relativism		531
	Introduction	533
41	Epistemological Realism Michael Williams	536
42	Justification, Meta-Epistemology, and Meaning Paul Moser	556
43	Reflective Equilibrium, Analytic Epistemology, and the Problem of Cognitive Diversity Stephen Stich	571
	Index	584

Preface

Epistemology is a philosophical inquiry into the nature, conditions, and extent of human knowledge. It encompasses some of the most puzzling and persistent issues in all of philosophy, ones that extensively define its history. The problem of skepticism is one example, and the empiricism/rationalism controversy another, along with its Kantian and Hegelian aftermath. Such issues, although alien to common sense at first sight, in fact derive naturally from straightforward reflection on the most ordinary knowledge about the world around us, knowledge produced or sustained through perception, memory, or induction. Elementary reflection on such matters produces puzzles and paradoxes that have engaged philosophers from ancient times to the present.

This anthology, meant to supplement Blackwell's *Companion to Epistemology and Guide to Epistemology*, collects selections representative of the best current discussion on the most central issues in the field. Although such work is inevitably demanding, our selections, some of which are only excerpts, should all prove accessible in proper order to the attentive reader who approaches these issues for the first time.

The selections are collected in ten sections, each of which opens with an introduction that discusses the contained readings, and is followed by a list of further readings on the subject matter of that section. For further expert but introductory discus-

sion of the issues, the reader is referred to the relevant Blackwell *Companion and Guide*.

The topics taken up in these ten sections by no means exhaust the field of epistemology. Space limits have made it impossible to include all topics in the field. We have consciously selected central issues. But we have also drawn from contemporary work some most novel and radical responses to these issues. The resulting collection brings together a variety of approaches and solutions, still under vigorous debate. We have not included work on more specific issues concerning memory, perception, other minds, induction, testimony, etc. On these issues excellent work has been published in recent decades and continues to appear. An important new development is feminist epistemology, but this has been extensively covered in another recent Blackwell publication, Linda Alcoff's *Epistemology: The Big Questions*.

Michael DiRamio and Brie Gertler helped early and variously, as did Joseph Shieber and Baron Reed, more recently. Matthew McGrath contributed the introductions to the individual sections, which accounts for the special acknowledgment on the title page. Also helpful were the comments and suggestions by anonymous referees. Steve Smith, our editor, has been unfailingly supportive and helpful, as have Sarah Dancy and Beth Remmes of the Blackwell editorial staff. Our thanks to them all!

Jaegwon Kim
Ernest Sosa

Acknowledgments

The authors and publishers would like to thank the following for permission to use copyright material:

American Philosophical Quarterly for chapters 14 (Appendix) and 21;

Analysis for chapter 7;

Australasian Journal of Philosophy for chapters 39 and 40;

Blackwell Publishers for chapters 15 and 41;

Bobbs-Merrill for chapter 13;

Cambridge University Press for chapters 26 and 36;

Canadian Journal of Philosophy for chapter 37;

Collier Books for chapters 2, 3 and 4;

Columbia University Press for chapters 5 and 23;

Cornell University Press for chapters 20 (with James Van Cleve) and 38 (with Keith DeRose),

Harvard University Press for chapters 10 and 25;

Kluwer Academic Publishers for chapters 16, 17, 27, 29, and 43;

MIT Press for chapter 34;

Oxford University Press for chapters 1, 35, and 42;

Philosophical Exchange for chapter 6;

Philosophical Topics for chapter 28 (with William P. Alston);

Philosophy and Phenomenological Research for chapters 32 and 33;

Prentice-Hall for chapter 11;

Princeton University Press for chapter 9;

Ridgeview Publishing Co. for chapter 24;

Rowman and Littlefield for chapters 18 and 31,

The Journal of Philosophy for chapters 8 and 22;

University of Minnesota Press for chapters 12 and 14;

Wadsworth for chapter 19;

Westview Press for chapter 30.

Every effort has been made to trace the copyright holders, but if any has been inadvertently overlooked, the publishers will be pleased to make the necessary arrangements at the first opportunity.



PART I

Skepticism

Introduction

Like René Descartes, we have all asked ourselves at one time or another “Couldn’t everything I seem to see, hear, etc. be illusory? Might I in fact be dreaming all this? If so, what do I really know of the outside world?” The skeptic’s answers are pessimistic: yes, you could be dreaming, and so you know nothing of the outside world. The conclusion is outlandish and yet the reasoning behind it hardly seems strained at all. We instinctively feel the pressure towards skepticism in the movement from the question about the trustworthiness of our senses to the question of our knowledge. Given that our knowledge of the outside world, or at least the bulk of it, derives from the senses, how can we know anything about the world unless we first show that our senses can be trusted? The core of skeptical strategy is in fact more general: how can one gain knowledge using a source of belief unless one first shows that the source is trustworthy?

In his selection, Barry Stroud presents the skeptic’s argument in its most favorable light. The skeptic does not hold us up to an uncommonly high standard of knowledge only to make the obvious point that we fail to meet it. The skeptic invokes only the standards presupposed in everyday knowledge attributions. To use an example of Stroud’s, if no goldfinch could possibly be a canary, then if one is to know that the bird one sees is a goldfinch, one must be able to rule out its being a canary. More generally, to know that *p*, one must be able to rule out every possibility one knows to be incompatible with one’s knowing that *p*. The skeptic then has her wedge: to know that you’re sitting beside a warm fire, you must be able to rule out any possibility which excludes this knowledge, including innumerable “skeptical possibilities,”

such as that you’re dreaming, that you’re being deceived by a malicious demon, and that you’re a brain in a vat stimulated to have the experiences and apparent memories you now have.

In each of the selections from the work of G. E. Moore, the tables are turned on the skeptic. Moore provides a counter-argument in “Proof of an External World.” A good proof, he explains, proceeds from known premises to a distinct conclusion to which they can be seen to lead. He then produces an example: raising his hands, one after the other, he exclaims “Here is a hand,” “Here is another hand,” and concludes “There are hands.” If asked to prove his premises, he would reject the demand, for not everything that is known can be proved.

Moore nevertheless takes the skeptic seriously. In “Certainty,” he grants that if he doesn’t know he is not dreaming, he doesn’t know he is standing up giving a lecture. Still he asks why there is any more plausibility in using this premise as part of a *modus ponens* argument to conclude that he doesn’t know he is standing up than in using it as a part of the corresponding *modus tollens* argument to conclude that he does know after all that he is not dreaming.

In “Four Forms of Scepticism,” Moore fully admits that skeptical scenarios are logically possible, but finds it more certain that something has gone awol in the skeptical argument than that he lacks knowledge that he has hands (or is holding a pencil). Moreover, he concludes that since the only way he *could* know this is through some inductive or analogical argument from the character of his experience, therefore such an argument *must* exist.

Introduction

Peter Strawson disputes Stroud's assumption that the only possible ways to defuse skepticism are to directly refute it, as Moore attempts to do, or to indirectly refute it by showing that it is unintelligible or self-defeating. There is a further alternative, viz. naturalism. Strawson's naturalism is not the naturalized epistemology of Quine, but is closer in spirit to the naturalisms of Hume and Wittgenstein. Hume claimed: "Tis vain to ask, whether there be body or not." Strawson agrees. The best answer to skepticism is that it is idle. We simply must take ourselves to know about the world. We literally cannot believe skepticism, just as we literally cannot believe there are no physical objects, nor that there are no other minds, nor that inductive arguments are no guide to the truth. But Strawson's naturalism isn't merely the individualistic psychological naturalism of Hume. As Wittgenstein argued, our social existence makes certain beliefs necessary for us. Moreover, which beliefs are necessary at which times depends on the context. For Wittgenstein, any context determines a group of beliefs as unquestionable, as beyond being justified or being unjustified, as part of the scaffolding of our view of ourselves and the world. Nonetheless, the beliefs that compose the scaffolding may vary from context to context, as different portions of the water in a flowing river are at rest alongside the bank or at the river bank at different times and different places. It is interesting to note that while Hume clearly recognizes the intelligibility of skeptical challenges, Wittgenstein arguably does not. Wittgenstein, in fact, may be more closely aligned with the second alternative Stroud discusses, that of rejecting skepticism as unintelligible.

Peter Unger provides a somewhat different argument for skepticism. He claims that if one knows that *p*, then one has the right to be absolutely certain that *p* and so to disregard evidence

pointing against *p*. A statement such as "He knew it was raining, but he wasn't absolutely certain that it was" is not only infelicitous, it is inconsistent. Unger then argues that no one ever has such a right, for this would be a right to be completely dogmatic, a right to admit no evidence as relevant and possibly damaging.

Of course, Unger must explain away our intuitions to the contrary in the hard cases, e.g., the case of knowing that you now exist. He describes a case in which a voice tells you that all of your experiences through your entire life, including the present experience of the voice, have been artificially induced by a team of scientists; in fact there are no trees, fields, etc. To establish its own credibility, the voice predicts you will have all varieties of unusual experiences; and upon the prediction, it becomes so. Now having gone through all this, is it reasonable for you to hold with absolute certainty that there are trees, fields, etc? Surely not. Now suppose the voice tells you that you often confuse the meaning of "to exist" with that of "to persist." "To persist" means *to be*, and "to exist" means *to continue on in the face of obstacles*. The voice also tells you that you are a changeable person, one who does not continue on in the face of obstacles. So, many of the times in the past in which you asserted confidently "I exist," you were wrong. Perhaps you will say: "well, I now persist, then!" Unger allows you may say this, but claims you should not be certain of its truth. For you could now be using "to persist" to mean *to continue on in the face of obstacles*. Unger acknowledges the difference between the fact that the sentence "I now exist" expresses a truth and the fact of one's own present existence. Still, he claims, you might be confident, on the voice's assurance, that "I now exist" as used by you expresses a truth, but still rationally be less than certain that you now exist.

Further Reading

Annas, Julia and Jonathan Barnes, *The Modes of Scepticism* (Cambridge: Cambridge University Press, 1985).
Burnyeat, M. F. (ed.), *The Skeptical Tradition* (Berkeley: University of California Press, 1983).
Clarke, Thompson, "The Legacy of Skepticism," *Journal of Philosophy* 69 (1972), pp. 754–69.
DeRose, K. and Warfield, T. (eds), *Skepticism: A Contemporary Reader* (Oxford: Oxford University Press, 1999).

Fumerton, R., *Metaepistemology and Skepticism* (Lanham, MD: Roman & Littlefield, 1995).
Hadot, Pierre, *Philosophy as a Way of Life* (Oxford: Blackwell, 1995).
Hankinson, R. J., *The Sceptics* (London: Routledge, 1995).
Klein, P., *Certainty: A Refutation of Scepticism* (Minneapolis: University of Minnesota Press, 1991).

- Nozick, Robert, *Philosophical Explanations* (Oxford: Oxford University Press, 1981).
- Popkin, Richard, *Scepticism from Erasmus to Spinoza* (Berkeley: University of California Press, 1979).
- Roth, Michael D. and Ross, Glenn (eds), *Doubling: Contemporary Perspectives on Skepticism* (Dordrecht: Kluwer Academic Publishers, 1990).
- Sosa, Ernest, "The Skeptic's Appeal Denied," in Marjorie Clay and Keith Lehrer (eds), *Knowledge and Skepticism* (Boulder: Westview Press, 1989).
- Strawson, P. F., *Skepticism and Naturalism: Some Varieties* (London: Methuen, 1985).
- Stroud, Barry, "Understanding Human Knowledge in General," in Marjorie Clay and Keith Lehrer (eds), *Knowledge and Skepticism* (Boulder: Westview Press, 1989), pp. 31–50.
- , *The Significance of Philosophical Skepticism* (Oxford: Oxford University Press, 1984).
- Unger, Peter, *Ignorance: A Case for Skepticism* (Oxford: Clarendon Press, 1975).
- Williams, Michael, *Unnatural Doubts* (Princeton: Princeton University Press, 1996).

The Problem of the External World

Barry Stroud

Since at least the time of Descartes in the seventeenth century there has been a philosophical problem about our knowledge of the world around us.¹ Put most simply, the problem is to show how we can have any knowledge of the world at all. The conclusion that we cannot, that no one knows anything about the world around us, is what I call 'scepticism about the external world', so we could also say that the problem is to show how or why scepticism about the external world is not correct. My aim is not to solve the problem but to understand it: I believe the problem has no solution; or rather that the only answer to the question as it is meant to be understood is that we can know nothing about the world around us. But how is the question meant to be understood? It can be expressed in a few English words familiar to all of us, but I hope to show that an understanding of the special philosophical character of the question, and of the inevitability of an unsatisfactory answer to it, cannot be guaranteed by our understanding of those words alone. To see how the problem is meant to be understood we must therefore examine what is perhaps best described as its source — how the problem arises and how it acquires that special character that makes an unsatisfactory negative answer inevitable. We must try to understand the *philosophical* problem of our knowledge of the external world.

The problem arose for Descartes in the course of reflecting on everything he knows. He reached a point in his life at which he tried to sit back and

reflect on everything he had ever been taught or told, everything he had learned or discovered or believed since he was old enough to know or believe anything.² We might say that he was reflecting on his knowledge, but putting it that way could suggest that what he was directing his attention to was indeed knowledge, and whether it was knowledge or not is precisely what he wanted to determine. 'Among all the things I believe or take to be true, what amounts to knowledge and what does not?'; that is the question Descartes asks himself. It is obviously a very general question, since it asks about everything he believes or takes to be true, but in other respects it sounds just like the sort of question we are perfectly familiar with in everyday life and often know how to answer.

For example, I have come to accept over the years a great many things about the common cold. I have always been told that one can catch cold by getting wet feet, or from sitting in a draught, or from not drying one's hair before going outdoors in cold weather. I have also learned that the common cold is the effect of a virus transmitted by an already infected person. And I also believe that one is more vulnerable to colds when over-tired, under stress, or otherwise in less than the best of health. Some of these beliefs seem to me on reflection to be inconsistent with some others; I see that it is very unlikely that all of them could be true. Perhaps they could be, but I acknowledge that there is much I do not understand. If I sit back and try to think about all my 'knowledge' of the common cold, then, I might easily come to wonder how much of it really amounts to knowledge and how much does not. What do I really know about the common cold? If I were sufficiently interested

Originally published in B. Stroud, *The Significance of Philosophical Scepticism* (Oxford: Clarendon Press, 1984), ch. 1.

in pursuing the matter it would be natural to look into the source of my beliefs. Has there ever been any good reason for thinking that colds are even correlated with wet hair in cold weather, for example, or with sitting in a draught? Are the people from whom I learned such things likely to have believed them for good reasons? Are those beliefs just old wives' tales, or are they really true, and perhaps even known to be true by some people? These are questions I might ask myself, and I have at least a general idea of how to go about answering them.

Apart from my impression of the implausibility of all my beliefs about the common cold being true together, I have not mentioned any other reason for being interested in investigating the state of my knowledge on that subject. But for the moment that does not seem to affect the intelligibility or the feasibility of the reflective project. There is nothing mysterious about it. It is the sort of task we can be led to undertake for a number of reasons, and often very good reasons, in so far as we have very good reasons for preferring knowledge and firm belief to guesswork or wishful thinking or simply taking things for granted.

Reflection on or investigation of our putative knowledge need not always extend to a wide area of interest. It might be important to ask whether some quite specific and particular thing I believe or have been taking for granted is really something I know. As a member of a jury I might find that I have been ruling out one suspect in my mind because he was a thousand miles away, in Cleveland, at the time of the crime. But I might then begin to ask myself whether that is really something that I know. I would reflect on the source of my belief, but reflection in this case need not involve a general scrutiny of everything I take myself to know about the case. Re-examining the man's alibi and the credentials of its supporting witnesses might be enough to satisfy me. Indeed I might find that its reliability on those counts is precisely what I had been going on all along.

In pointing out that we are perfectly familiar with the idea of investigating or reviewing our knowledge on some particular matter or in some general area I do not mean to suggest that it is always easy to settle the question. Depending on the nature of the case, it might be very difficult, perhaps even impossible at the time, to reach a firm conclusion. For example, it would probably be very difficult if not impossible for me to trace and assess the origins of many of those things I

believe about the common cold. But it is equally true that sometimes it is not impossible or even especially difficult to answer the question. We do sometimes discover that we do not really know what we previously thought we knew. I might find that what I had previously believed is not even true – that sitting in draughts is not even correlated with catching a cold, for example. Or I might find that there is not or perhaps never was any good reason to believe what I believed – that the man's alibi was concocted and then falsely testified to by his friends. I could reasonably conclude in each case that I, and everyone else for that matter, never did know what I had previously thought I knew. We are all familiar with the ordinary activity of reviewing our knowledge, and with the experience of reaching a positive verdict in some cases and a negative verdict in others.

Descartes's own interest in what he knows and how he knows it is part of his search for what he calls a general method for 'rightly conducting reason and seeking truth in the sciences'.³ He wants a method of inquiry that he can be assured in advance will lead only to the truth if properly followed. I think we do not need to endorse the wisdom of that search or the feasibility of that programme in order to try to go along with Descartes in his general assessment of the position he is in with respect to the things he believes. He comes to find his putative knowledge wanting in certain general respects, and it is in the course of that original negative assessment that the problem I am interested in arises. I call the assessment 'negative' because by the end of his *First Meditation* Descartes finds that he has no good reason to believe anything about the world around him and therefore that he can know nothing of the external world.

How is that assessment conducted, and how closely does it parallel the familiar kind of review of our knowledge that we all know how to conduct in everyday life? The question in one form or another will be with us for the rest of this book. It is the question of what exactly the problem of our knowledge of the external world amounts to, and how it arises with its special philosophical character. The source of the problem is to be found somewhere within or behind the kind of thinking Descartes engages in.

One way Descartes's question about his knowledge differs from the everyday examples I considered is in being concerned with *everything* he believes or takes to be true. How does one go about

assessing all of one's knowledge all at once? I was able to list a few of the things I believe about the common cold and then to ask about each of them whether I really know it, and if so how. But although I can certainly list a number of the things I believe, and I would assent to many more of them as soon as they were put to me, there obviously is no hope of assessing everything I believe in this piecemeal way. For one thing, it probably makes no sense, strictly speaking, to talk of the number of things one believes. If I am asked whether it is one of my beliefs that I went to see a film last night I can truly answer 'Yes'. If I were asked whether it is one of my beliefs that I went to the movies last night I would give the same answer. Have I thereby identified two, or only one, of my beliefs? How is that question ever to be settled? If we say that I identified only one of my beliefs, it would seem that I must also be said to hold the further belief that going to see a film and going to the movies are one and the same thing. So we would have more than one belief after all. The prospects of arriving even at a principle for counting beliefs, let alone at an actual number of them, seem dim.

Even if it did make sense to count the things we believe it is pretty clear that the number would be indefinitely large and so an assessment of our beliefs one by one could never be completed anyway. This is easily seen by considering only some of the simplest things one knows, for example in arithmetic. One thing I know is that one plus one equals two. Another thing I know is that one plus two is three, and another, that one plus three is four. Obviously there could be no end to the task of assessing my knowledge if I had to investigate separately the source of each one of my beliefs in that series. And even if I succeeded I would only have assessed the things I know about the addition of the number one to a given number; I would still have to do the same for the addition of two, and then the addition of three, and so on. And even that would exhaust only my beliefs about addition; all my other mathematical beliefs, not to mention all the rest of my knowledge, would remain so far unexamined. Obviously the job cannot be done piecemeal, one by one. Some method must be found for assessing large classes of beliefs all at once.

One way to do this would be to look for common sources or channels or bases of our beliefs, and then to examine the reliability of those sources or bases, just as I examined the source or basis of my belief that the suspect was in Cleveland. Des-

cartes describes such a search as a search for 'principles' of human knowledge, 'principles' whose general credentials he can then investigate (HR, 145). If some 'principles' are found to be involved in all or even most of our knowledge, an assessment of the reliability of those 'principles' could be an assessment of all or most of our knowledge. If I found good reason to doubt the reliability of the suspect's alibi, for example, and that was all I had to go on in my belief that he was in Cleveland, then what I earlier took to be my knowledge that he was in Cleveland would have been found wanting or called into question. Its source or basis would have been undermined. Similarly, if one of the 'principles' or bases on which all my knowledge of the world depends were found to be unreliable, my knowledge of the world would to that extent have been found wanting or called into question as well.

Are there any important 'principles' of human knowledge in Descartes's sense? It takes very little reflection on the human organism to convince us of the importance of the senses – sight, hearing, touch, taste, and smell. Descartes puts the point most strongly when he says that 'all that up to the present time I have accepted as most true and certain I have learned either from the senses or through the senses' (HR, 145). Exactly what he would include under 'the senses' here is perhaps somewhat indeterminate, but even if it is left vague many philosophers would deny what Descartes appears to be saying. They would hold that, for example, the mathematical knowledge I mentioned earlier is not and could not be acquired from the senses or through the senses, so not *everything* I know is known in that way. Whether Descartes is really denying the views of those who believe in the non-sensory character of mathematical knowledge, and whether, if he were, he would be right, are issues we can set aside for the moment. It is clear that the senses are at least very important for human knowledge. Even restricting ourselves to the traditional five senses we can begin to appreciate their importance by reflecting on how little someone would ever come to know without them. A person blind and deaf from birth who also lacked taste buds and a sense of smell would know very little about anything, no matter how long he lived. To imagine him also anaesthetized or without a sense of touch is perhaps to stretch altogether too far one's conception of a human organism, or at least a human organism from whom we can hope to learn something about human knowledge. The importance of the senses as a source or channel of

knowledge seems undeniable. It seems possible, then, to acknowledge their importance and to assess the reliability of that source, quite independently of the difficult question of whether *all* our knowledge comes to us in that way. We would then be assessing the credentials of what is often called our 'sensory' or 'experiential' or 'empirical' knowledge, and that, as we shall see, is quite enough to be going on with.

Having found an extremely important 'principle' or source of our knowledge, how can we investigate or assess *all* the knowledge we get from that source? As before, we are faced with the problem of the inexhaustibility of the things we believe on that basis, so no piecemeal, one-by-one procedure will do. But perhaps we can make a sweeping negative assessment. It might seem that as soon as we have found that the senses are one of the sources of our beliefs we are immediately in a position to condemn all putative knowledge derived from them. Some philosophers appear to have reasoned in this way, and many have even supposed that Descartes is among them. The idea is that if I am assessing the reliability of my beliefs and asking whether I really know what I take myself to know, and I come across a large class of beliefs which have come to me through the senses, I can immediately dismiss all those beliefs as unreliable or as not amounting to knowledge because of the obvious fact that I can sometimes be wrong in my beliefs based on the senses. Things are not always as they appear, so if on the basis of the way they appear to me I believe that they really are a certain way, I might still be wrong. We have all found at one time or another that we have been misled by appearances; we know that the senses are not always reliable. Should we not conclude, then, that as a general source of knowledge the senses are not to be trusted? As Descartes puts it, is it not wiser never 'to trust entirely to any thing by which we have once been deceived' (HR, 145)? Don't we have here a quite general way of condemning as not fully reliable *all* of our beliefs acquired by means of the senses?

I think the answer to that question is 'No, we do not', and I think Descartes would agree with that answer. It is true that he does talk of the senses 'deceiving' us on particular occasions, and he does ask whether that is not enough to condemn the senses in general as a source of knowledge, but he immediately reminds us of the obvious fact that the circumstances in which the senses 'deceive' us might be special in certain ascertainable ways, and

so their occasional failures would not support a blanket condemnation of their reliability.

Sometimes, to give an ancient example, a tower looks round from a distance when it is actually square. If we relied only on the appearances of the moment we might say that the distant tower is round, and we would be wrong. We also know that there are many small organisms invisible to the naked eye. If the table before me is covered with such organisms at the moment but I look at it and say there is nothing on the table at all, once again I will be wrong. But all that follows from these familiar facts, as Descartes points out, is that there are things about which we can be wrong, or there are situations in which we can get false beliefs, if we rely entirely on our senses at that moment. So sometimes we should be careful about what we believe on the basis of the senses, or sometimes perhaps we should withhold our assent from any statement about how things are – when things are too far away to be seen properly, for example, or too small to be seen at all. But that obviously is not enough to support the policy of never trusting one's senses, or never believing anything based on them. Nor does it show that I can never know anything by means of the senses. If my car starts promptly every morning for two years in temperate weather at sea level but then fails to start one morning in freezing weather at the top of a high mountain, that does not support the policy of never trusting my car to start again once I return to the temperate lower altitude from which I so foolishly took it. Nor does it show that I can never know whether my car will ever start again. It shows only that there are certain circumstances in which my otherwise fully reliable car might not start. So the fact that we are sometimes wrong or 'deceived' in our judgements based on the senses is not enough in itself to show that the senses are never to be trusted and are therefore never reliable as a source of knowledge.

Descartes's negative assessment of all of his sensory knowledge does not depend on any such reasoning. He starts his investigation, rather, in what would seem to be the most favourable conditions for the reliable operation of the senses as a source of knowledge. While engaging in the very philosophical reflections he is writing about in his *First Meditation* Descartes is sitting in a warm room, by the fire, in a dressing-gown, with a piece of paper in his hand. He finds that although he might be able to doubt that a distant tower that looks round really is round, it seems impossible to

doubt that he really is sitting there by the fire in his dressing gown with a piece of paper in his hand. The fire and the piece of paper are not too small or too far away to be seen properly, they are right there before his eyes; it seems to be the best kind of position someone could be in for getting reliable beliefs or knowledge by means of the senses about what is going on around him. That is just how Descartes regards it. Its being a best-possible case of that kind is precisely what he thinks enables him to investigate or assess at one fell swoop all our sensory knowledge of the world around us. The verdict he arrives at about his putative knowledge that he is sitting by the fire with a piece of paper in his hand in that particular situation serves as the basis for a completely general assessment of the senses as a source of knowledge about the world around us.

How can that be so? How can he so easily reach a general verdict about all his sensory knowledge on the basis of a single example? Obviously not simply by generalizing from one particular example to all cases of sensory knowledge, as one might wildly (leap) to a conclusion about all red-haired men on the basis of one or two individuals. Rather, he takes the particular example of his conviction that he is sitting by the fire with a piece of paper in his hand as representative of the best position any of us can ever be in for knowing things about the world around us on the basis of the senses. What is true of a representative case, if it is truly representative and does not depend on special peculiarities of its own, can legitimately support a general conclusion. A demonstration that a particular isosceles triangle has a certain property, for example, can be taken as a demonstration that all isosceles triangles have that property, as long as the original instance was typical or representative of the whole class. Whether Descartes's investigation of the general reliability of the senses really does follow that familiar pattern is a difficult question. Whether, or in precisely what sense, the example he considers can be treated as representative of our relation to the world around us is, I believe, the key to understanding the problem of our knowledge of the external world. But if it turns out that there is nothing illegitimate about the way his negative conclusion is reached, the problem will be properly posed.

For the moment I think at least this much can be said about Descartes's reasoning. He chooses the situation in which he finds himself as representative of the best position we can be in for

knowing things about the world in the sense that, if it is impossible for him in that position to know that he is sitting by the fire with a piece of paper in his hand then it is also impossible for him in other situations to know anything about the world around him on the basis of his senses. A negative verdict in the chosen case would support a negative verdict everywhere else. The example Descartes considers is in that sense meant to be the *best* kind of case there could be of sensory knowledge about the world around us. I think we must admit that it is very difficult to see how Descartes or anyone else could be any better off with respect to knowing something about the world around him on the basis of the senses than he is in the case he considers. But if no one could be in any better position for knowing, it seems natural to conclude that any negative verdict arrived at about this example, any discovery that Descartes's beliefs in this case are not reliable or do not amount to knowledge, could safely be generalized into a negative conclusion about all of our sensory 'knowledge' of the world. If candidates with the best possible credentials are found wanting, all those with less impressive credentials must fall short as well.

It will seem at first sight that in conceding that the whole question turns on whether Descartes knows in this particular case we are conceding very little; it seems obvious that Descartes on that occasion does know what he thinks he knows about the world around him. But in fact Descartes finds that he cannot know in this case that he is sitting by the fire with a piece of paper in his hand. If the case is truly representative of our sensory knowledge in general, that will show that no one can know anything about the world around us. But how could he ever arrive at that negative verdict in the particular case he considers? How could anyone possibly doubt in such a case that the fire and the piece of paper are there? The paper is in Descartes's hand, the fire is right there before his open eyes, and he feels its warmth. Wouldn't anyone have to be mad to deny that he can know something about what is going on around him in those circumstances? Descartes first answers 'Yes'. He says that if he were to doubt or deny on that occasion that he is sitting by the fire with a piece of paper in his hand he would be no less mad than those paupers who say they are kings or those madmen who think they are pumpkins or are made of glass. But his reflections continue:

At the same time I must remember that I am a man, and that consequently I am in the habit of sleeping, and in my dreams representing to myself the same things or sometimes even less probable things, than do those who are insane in their waking moments. How often has it happened to me that in the night I dreamt that I found myself in this particular place, that I was dressed and seated near the fire, whilst in reality I was lying undressed in bed! At this moment it does indeed seem to me that it is with eyes awake that I am looking at this paper; that this head which I move is not asleep, that it is deliberately and of set purpose that I extend my hand and perceive it; what happens in sleep does not appear so clear nor so distinct as does all this. But in thinking over this I remind myself that on many occasions I have in sleep been deceived by similar illusions, and in dwelling carefully on this reflection I see so manifestly that there are no certain indications by which we may clearly distinguish wakefulness from sleep that I am lost in astonishment. And my astonishment is such that it is almost capable of persuading me that I now dream. (HR, 145–6)

With this thought, if he is right, Descartes has lost the whole world. He knows what he is experiencing, he knows how things appear to him, but he does not know whether he is in fact sitting by the fire with a piece of paper in his hand. It is, for him, exactly as if he were sitting by the fire with a piece of paper in his hand, but he does not know whether there really is a fire or a piece of paper there or not; he does not know what is really happening in the world around him. He realizes that if everything he can ever learn about what is happening in the world around him comes to him through the senses, but he cannot tell by means of the senses whether or not he is dreaming, then all the sensory experiences he is having are compatible with his merely dreaming of a world around him while in fact that world is very different from the way he takes it to be. That is why he thinks he must find some way to tell that he is not dreaming. Far from its being mad to deny that he knows in this case, he thinks his recognition of the possibility that he might be dreaming gives him 'very powerful and maturely considered' (HR, 148) reasons for withholding his judgement about how things are in the world around him. He thinks it is eminently reasonable to insist that if he is to

know that he is sitting by the fire he must know that he is not dreaming that he is sitting by the fire. That is seen as a necessary condition of knowing something about the world around him. And he finds that that condition cannot be fulfilled. On careful reflection he discovers that 'there are no certain indications by which we may clearly distinguish wakefulness from sleep'. He concludes that he knows nothing about the world around him because he cannot tell that he is not dreaming; he cannot fulfil one of the conditions necessary for knowing something about the world.

The Cartesian problem of our knowledge of the external world therefore becomes: how can we know anything about the world around us on the basis of the senses if the senses give us only what Descartes says they give us? What we gain through the senses is on Descartes's view only information that is compatible with our dreaming things about the world around us and not knowing anything about that world. How then can we know anything about the world by means of the senses? The Cartesian argument presents a challenge to our knowledge, and the problem of our knowledge of the external world is to show how that challenge can be met.

When I speak here of the Cartesian argument or of Descartes's sceptical conclusion or of his negative verdict about his knowledge I refer of course only to the position he finds himself in by the end of his *First Meditation*. Having at that point discovered and stated the problem of the external world, Descartes goes on in the rest of his *Meditations* to try to solve it, and by the end of the *Sixth Meditation* he thinks he has explained how he knows almost all those familiar things he began by putting in question. So when I ascribe to Descartes the view that we can know nothing about the world around us I do not mean to suggest that that is his final and considered view; it is nothing more than a conclusion he feels almost inevitably driven to at the early stages of his reflections. But those are the only stages of his thinking I am interested in here. That is where the philosophical problem of our knowledge of the external world gets posed, and before we can consider possible solutions we must be sure we understand exactly what the problem is.

I have described it as that of showing or explaining how knowledge of the world around us is possible by means of the senses. It is important to keep in mind that that demand for an explanation arises in the face of a challenge or apparent

obstacle to our knowledge of the world. The possibility that he is dreaming is seen as an obstacle to Descartes's knowing that he is sitting by the fire, and it must be explained how that obstacle can either be avoided or overcome. It must be shown or explained *how* it is possible for us to know things about the world, given that the sense-experiences we get are compatible with our merely dreaming. Explaining how something is nevertheless possible, despite what looks like an obstacle to it, requires more than showing merely that there is no impossibility involved in the thing – that it is consistent with the principles of logic and the laws of nature and so in that sense *could* exist. The mere possibility of the state of affairs is not enough to settle the question of how our knowledge of the world is possible; we must understand how the apparent obstacle is to be got round.

Descartes's reasoning can be examined and criticized at many different points, and has been closely scrutinized by many philosophers for centuries. It has also been accepted by many, perhaps by more than would admit or even realize that they accept it. There seems to me no doubt about the force and the fascination – I would say the almost overwhelming persuasiveness – of his reflections. That alone is something that needs accounting for. I cannot possibly do justice to all reasonable reactions to them here. In the rest of this chapter I want to concentrate on deepening and strengthening the problem and trying to locate more precisely the source of its power.

There are at least three distinct questions that could be pressed. Is the possibility that Descartes might be dreaming really a threat to his knowledge of the world around him? Is he right in thinking that he must know that he is not dreaming if he is to know something about the world around him? And is he right in his 'discovery' that he can never know that he is not dreaming? If Descartes were wrong on any of these points it might be possible to avoid the problem and perhaps even to explain without difficulty how we know things about the world around us.

On the first question, it certainly seems right to say that if Descartes were dreaming that he is sitting by the fire with a piece of paper in his hand he would not then know that he is sitting by the fire with a piece of paper in his hand. When you dream that something is going on in the world around you you do not thereby know that it is. Most often, of course, what we dream is not even true; no one is actually chasing us when we are

lying asleep in bed dreaming, nor are we actually climbing stairs. But although usually what we dream is not really so, that is not the real reason for our lack of knowledge. Even if Descartes were in fact sitting by the fire and actually had a piece of paper in his hand at the very time he was dreaming that he is sitting by the fire with a piece of paper in his hand, he would not thereby know he was sitting there with that paper. He would be like a certain Duke of Devonshire who, according to G. E. Moore, once dreamt he was speaking in the House of Lords and woke up to find that he *was* speaking in the House of Lords.⁴ What he was dreaming was in fact so. But even if what you are dreaming is in fact so you do not thereby know that it is. Even if we allow that when you are dreaming that something is so you can be said, at least for the time being, to think or to believe that it is so, there is still no real connection between your thinking or believing what you do and its being so. At best you have a thought or a belief which just happens to be true, but that is no more than coincidence and not knowledge. So Descartes's first step relies on what seems to be an undeniable fact about dreams: if you are dreaming that something is so you do not thereby know that it is so.

This bald claim needs to be qualified and more carefully explained, but I do not think that will diminish the force of the point for Descartes's purposes. Sometimes what is going on in the world around us has an effect on what we dream; for example, a banging shutter might actually cause me to dream, among other things, that a shutter is banging. If my environment affects me in that way, and if in dreams I can be said to think or believe that something is so, would I not in that case know that a shutter is banging? It seems to me that I would not, but I confess it is difficult to say exactly why I think so. That is probably because it is difficult to say exactly what is required for knowledge. We use the term 'know' confidently, we quite easily distinguish cases of knowledge from cases of its absence, but we are not always in a position to state what we are going on in applying or withholding the term in the ways we do. I think that in the case of the banging shutter it would not be knowledge because I would be *dreaming*, I would not even be awake. At least it can be said, I think, that even if Descartes's sitting by the fire with a piece of paper in his hand (like the banging shutter) is what in fact causes him to dream that he is sitting by the fire with a piece of

paper in his hand, that is still no help to him in coming to know what is going on in the world around him. He realizes that he could be dreaming that he is sitting by the fire even if he is in fact sitting there, and that is the possibility he finds he has to rule out.

I have said that if you are dreaming that something is so you do not thereby know that it is so, and it might seem as if that is not always true. Suppose a man and a child are both sleeping. I say of the child that it is so young it does not know what seven times nine is, whereas the grown man does know that. If the man happens at that very moment to be dreaming that seven times nine is sixty-three (perhaps he is dreaming that he is computing his income tax), then he is a man who is dreaming that something is so and also knows that it is so. The same kind of thing is possible for knowledge about the world around him. He might be a physicist who knows a great deal about the way things are which the child does not know. If the man also dreams that things are that way he can once again be said to be dreaming that something is so and also to know that it is so. There is therefore no incompatibility between dreaming and knowing. That is true, but I do not think it affects Descartes's argument. He is led to consider how he knows he is not dreaming at the moment by reflecting on how he knows at that moment that he is sitting by the fire with a piece of paper in his hand. If he knows that at all, he thinks, he knows it on the basis of the senses. But he realizes that his having the sensory experiences he is now having is compatible with his merely dreaming that he is sitting by the fire with a piece of paper in his hand. So he does not know on the basis of the sensory experiences he is having at the moment that he is sitting by the fire. Nor, of course, did the man in my examples know the things he was said to know on the basis of the sensory experiences he was having at that moment. He knew certain things to be so, and he was dreaming those things to be so, but in dreaming them he did not *thereby* know them to be so.

But as long as we allow that the sleeping man does know certain things about the world around him, even if he does not know them on the basis of the very dreams he is having at the moment, isn't that enough to show that Descartes must nevertheless be wrong in his conclusion that no one can know anything about the world around him? No. It shows at most that we were hasty or were ignoring Descartes's conclusion in conceding that someone

could know something about the world around him. If Descartes's reasoning is correct the dreaming physicist, even when he is awake, does not really know any of the things we were uncritically crediting him with knowing about the way things are – or at least he does not know them on the basis of the senses. In order to know them on the basis of the senses there would have to have been at least some time at which he knew something about what was going on around him at that time. But if Descartes is right he could not have known any such thing unless he had established that he was not dreaming at that time; and according to Descartes he could never establish that. So the fact about dreams that Descartes relies on – that one who dreams that something is so does not thereby know that it is so – is enough to yield his conclusion if the other steps of his reasoning are correct.

When he first introduces the possibility that he might be dreaming Descartes seems to be relying on some knowledge about how things are or were in the world around him. He says 'I remind myself that on many occasions I have in sleep been deceived by similar illusions', so he seems to be relying on some knowledge to the effect that he has actually dreamt in the past and that he remembers having been 'deceived' by those dreams. That is more than he actually needs for his reflections about knowledge to have the force he thinks they have. He does not need to support his judgement that he has actually dreamt in the past. The only thought he needs is that it is now *possible* for him to be dreaming that he is sitting by the fire, and that if that possibility were realized he would not know that he is sitting by the fire. Of course it was no doubt true that Descartes had dreamt in the past and that his knowledge that he had done so was partly what he was going on in acknowledging the possibility of his dreaming on this particular occasion. But neither the fact of past dreams nor knowledge of their actual occurrence would seem to be strictly required in order to grant what Descartes relies on – the possibility of dreaming, and the absence of knowledge if that possibility were realized. The thought that he *might* be dreaming that he is sitting by the fire with a piece of paper in his hand, and the fact that if he were he wouldn't know he was sitting there, is what gives Descartes pause. That would worry him in the way it does even if he had never actually had any dreams exactly like it in the past – if he had never dreamt about fires and pieces of paper at all. In fact, I think he need never have actually dreamt of anything

before, and certainly needn't know that he ever has, in order to be worried in the way he is by the thought that he might be dreaming now.

The fact that the possibility of dreaming is all Descartes needs to appeal to brings out another truth about dreams that his argument depends on – that anything that can be going on or that one can experience in one's waking life can also be dreamt about. This again is only a statement of possibility – no sensible person would suggest that we *do* at some time dream of everything that actually happens to us, or that everything we dream about does in fact happen sometime. But it is very plausible to say that there is nothing we *could* not dream about, nothing that could be the case that we *could* not dream to be the case. I say it is very plausible; of course I cannot prove it to be true. But even if it is not true with complete generality, we must surely grant that it is possible to dream that one is sitting by a fire with a piece of paper in one's hand, and possible to dream of countless other equally obvious and equally mundane states of affairs as well, and those possibilities are what Descartes sees as threatening to his knowledge of the world around him.

There seems little hope, then, of objecting that it is simply not possible for Descartes to dream that he is sitting by the fire with a piece of paper in his hand. Nor is it any more promising to say that even if he were dreaming it would not follow that he did not know that he was sitting there. I think both those steps or assumptions of Descartes's reasoning are perfectly correct, and further defence of them at this stage is unnecessary. If his argument and the problem to which it gives rise are to be avoided, it might seem that the best hope is therefore to accept his challenge and show that it can be met. That would be in effect to argue that Descartes's alleged 'discovery' is no discovery at all: we *can* sometimes know that we are not dreaming.

This can easily seem to be the most straightforward and most promising strategy. It allows that Descartes is right in thinking that knowing that one is not dreaming is a condition of knowing something about the world around us, but wrong in thinking that that condition can never be met. And that certainly seems plausible. Surely it is not impossible for me to know that I am not dreaming? Isn't that something I often know, and isn't it something I can sometimes find out if the question arises? If it is, then the fact that I must know that I am not dreaming if I am to know anything about

the world around me will be no threat to my knowledge of the world.

However obvious and undeniable it might be that we often do know that we are not dreaming, I think this straightforward response to Descartes's challenge is a total failure. In calling it straightforward I mean that it accepts Descartes's conditions for knowledge of the world and tries to show that they can be fulfilled. That is what I think cannot be done. To put the same point in another way: I think Descartes would be perfectly correct in saying 'there are no certain indications by which we may clearly distinguish wakefulness from sleep', and so we could never tell we are not dreaming, *if* he were also right that knowing that one is not dreaming is a condition of knowing something about the world around us. That is why I think one cannot accept that condition and then go on to establish that one is not dreaming. I do not mean to be saying simply that Descartes is right – that we can never know that we are not dreaming. But I do want to argue that either we can never know that we are not dreaming or else what Descartes says is a condition of knowing things about the world is not really a condition in general of knowing things about the world. The straightforward strategy denies both alternatives. I will try to explain why I think we must accept one alternative or the other.

When Descartes asks himself how he knows that he is sitting by the fire with a piece of paper in his hand why does he immediately go on to ask himself how he knows he is not dreaming that he is sitting by the fire with a piece of paper in his hand? I have suggested that it is because he recognizes that if he were dreaming he would not know on the basis of his senses at the moment that he is sitting there, and so he thinks he must know that that possibility does not obtain if he is to know that he is in fact sitting there. But this particular example was chosen, not for any peculiarities it might be thought to possess, but because it could be taken as typical of the best position we can ever be in for coming to know things about the world around us on the basis of the senses. What is true of this case that is relevant to Descartes's investigation of knowledge is supposed to be true of all cases of knowledge of the world by means of the senses; that is why the verdict arrived at here can be taken to be true of our sensory knowledge generally. But what Descartes thinks is true of this particular case of sensory knowledge of the world is that he must know he is not dreaming if he is to know that he is

sitting by the fire with a piece of paper in his hand. That is required, not because of any peculiarities of this particular case, but presumably because, according to Descartes, it is a necessary condition of any case – even a best possible case – of knowledge of the world by means of the senses. That is why I ascribed to Descartes the quite general thesis that knowing that one is not dreaming is a condition of knowing something about the world around us on the basis of the senses. Since he thinks the possibility of his dreaming must be ruled out in the case he considers, and the case he considers is regarded as typical and without special characteristics of its own, he thinks that the possibility that he is dreaming must be ruled out in every case of knowing something about the world by means of the senses.

If that really is a condition of knowing something about the world, I think it can be shown that Descartes is right in holding that it can never be fulfilled. That is what the straightforward response denies, and that is why I think that response must be wrong. We cannot accept the terms of Descartes's challenge and then hope to meet it.

Suppose Descartes tries to determine that he is not dreaming in order to fulfil what he sees as a necessary condition of knowing that he is sitting by the fire with a piece of paper in his hand. How is he to proceed? He realizes that his seeing his hand and seeing and feeling a piece of paper before him and feeling the warmth of the fire – in fact his getting all the sensory experiences or all the sensory information he is then getting – is something that could be happening even if he were dreaming. To establish that he is not dreaming he would therefore need something more than just those experiences or that information alone. He would also need to know whether those experiences and that information are reliable, not merely dreamt. If he could find some operation or test, or if he could find some circumstance or state of affairs, that indicated to him that he was not dreaming, perhaps he could then fulfil the condition – he could know that he was not dreaming. But how could a test or a circumstance or a state of affairs indicate to him that he is not dreaming *if* a condition of knowing *anything* about the world is that he know he is not dreaming? It could not. He could never fulfil the condition.

Let us suppose that there is in fact some test which a person can perform successfully only if he is not dreaming, or some circumstance or state of affairs which obtains only if that person is not

dreaming. Of course for that test or state of affairs to be of any use to him Descartes would have to know of it. He would have to know that there is such a test or that there is a state of affairs that shows that he is not dreaming; without such information he would be no better off for telling that he is not dreaming than he would be if there were no such test or state of affairs at all. To have acquired that information he would at some time have to have known more than just something about the course of his sensory experience, since the connection between the performance of a certain test, or between a certain state of affairs, and someone's not dreaming is not itself just a fact about the course of that person's sensory experience; it is a fact about the world beyond his sensory experiences. Now strictly speaking if it is a condition of knowing *anything* about the world beyond one's sensory experiences that one know that one is not dreaming, there is an obvious obstacle to Descartes's ever having got the information he needs about that test or state of affairs. He would have to have known at some time that he was not dreaming in order to get the information he needs to tell at *any* time that he is not dreaming – and that cannot be done.

But suppose we forget about this difficulty and concede that Descartes does indeed know (somehow) that there is a test or circumstance or state of affairs that unfailingly indicates that he is not dreaming. Still, there is an obstacle to his ever using that test or state of affairs to tell that he is not dreaming and thereby fulfilling the condition for knowledge of the world. The test would have to be something he could know he had performed successfully, the state of affairs would have to be something he could know obtains. If he completely unwittingly happened to perform the test, or if the state of affairs happened to obtain but he didn't know that it did, he would be in no better position for telling whether he was dreaming than he would be if he had done nothing or did not even know that there was such a test. But how is he to know that the test has been performed successfully or that the state of affairs in question does in fact obtain? Anything one can experience in one's waking life can also be dreamt about; it is possible to dream that one has performed a certain test or dream that one has established that a certain state of affairs obtains. And, as we have seen, to dream that something about the world around you is so is not thereby to know that it is so. In order to know that his test has been performed or that the state of

affairs in question obtains Descartes would therefore have to establish that he is not merely dreaming that he performed the test successfully or that he established that the state of affairs obtains. How could that in turn be known? Obviously the particular test or state of affairs already in question cannot serve as a guarantee of its own authenticity, since it might have been merely dreamt, so some further test or state of affairs would be needed to indicate that the original test was actually performed and not merely dreamt, or that the state of affairs in question was actually ascertained to obtain and not just dreamt to obtain. But this further test or state of affairs is subject to the same general condition in turn. *Every* piece of knowledge that goes beyond one's sensory experiences requires that one know one is not dreaming. This second test or state of affairs will therefore be of use only if Descartes knows that he is not merely dreaming that he is performing or ascertaining it, since merely to dream that he had established the authenticity of the first test is not to have established it. And so on. At no point can he find a test for not dreaming which he can know has been successfully performed or a state of affairs correlated with not dreaming which he can know obtains. He can therefore never fulfil what Descartes says is a necessary condition of knowing something about the world around him. He can never know that he is not dreaming.

I must emphasize that this conclusion is reached *only* on the assumption that it is a condition of knowing anything about the world around us on the basis of the senses that we know we are not dreaming that the thing is so. I think it is his acceptance of that condition that leads Descartes to 'see so manifestly that there are no certain indications by which we may clearly distinguish wakefulness from sleep'. And I think Descartes is absolutely right to draw that conclusion, *given* what he thinks is a condition of knowledge of the world. But all I have argued on Descartes's behalf (he never spells out his reasoning) is that we cannot both accept that condition of knowledge and hope to fulfil it, as the straightforward response hopes to do. And of course if one of the necessary conditions of knowledge of the world can never be fulfilled, knowledge of the world around us will be impossible.

I think we have now located Descartes's reason for his negative verdict about sensory knowledge in general. If we agree that he must know that he is not dreaming if he is to know in his particular case

that he is sitting by the fire with a piece of paper in his hand, we must also agree that we can know nothing about the world around us.

Once we recognize that the condition Descartes takes as necessary can never be fulfilled if he is right in thinking it is indeed necessary, we are naturally led to the question whether Descartes is right. Is it really a condition of knowing something about the world that one know one is not dreaming? That is the second of the three questions I distinguished. It is the one that has received the least attention. In asking it now I do not mean to be going back on something I said earlier was undeniably true, viz., that if one is dreaming that something about the world is so *one does not* thereby know that it is so. That still seems to me undeniable, but it is not the same as Descartes's assumption that one must know that one is not dreaming if one is to know something about the world. The undeniable truth says only that you lack knowledge if you are dreaming; Descartes says that you lack knowledge if you don't know that you are not dreaming. Only with the stronger assumption can his sceptical conclusion be reached.

Is that assumption true? In so far as we find Descartes's reasoning convincing, or even plausible, I think it is because we too on reflection find that it is true. I said that not much attention had been paid to that particular part of Descartes's reasoning, and I think that too is because, as he presents it, the step seems perfectly convincing and so only other parts of the argument appear vulnerable. Why is that so? Is it because Descartes's assumption is indeed true? Is there anything we can do that would help us determine whether it is true or not? The question is important because I have argued so far that *if* it is true we can never know anything about the world around us on the basis of the senses, and philosophical scepticism about the external world is correct. We would have to find that conclusion as convincing or as plausible as we find the assumption from which it is derived.

Given our original favourable response to Descartes's reasoning, then, it can scarcely be denied that what I have called his assumption or condition *seems* perfectly natural to insist on. Perhaps it seems like nothing more than an instance of a familiar commonplace about knowledge. We are all aware that, even in the most ordinary circumstances when nothing very important turns on the outcome, we cannot know a particular thing unless

we have ruled out certain possibilities that we recognize are incompatible with our knowing that thing.

Suppose that on looking out the window I announce casually that there is a goldfinch in the garden. If I am asked how I know it is a goldfinch and I reply that it is yellow, we all recognize that in the normal case that is not enough for knowledge. 'For all you've said so far,' it might be replied, 'the thing could be a canary, so how do you know it's a goldfinch?' A certain possibility compatible with everything I have said so far has been raised, and if what I have said so far is all I have got to go on and I don't know that the thing in the garden is not a canary, then I do not know that there is a goldfinch in the garden. I must be able to rule out the possibility that it is a canary if I am to know that it is a goldfinch. Anyone who speaks about knowledge and understands what others say about it will recognize this fact or condition in particular cases.

In this example what is said to be possible is something incompatible with the truth of what I claim to know – if that bird were a canary it would not be a goldfinch in the garden, but a canary. What I believe in believing it is a goldfinch would be false. But that is not the only way a possibility can work against my knowledge. If I come to suspect that all the witnesses have conspired and made up a story about the man's being in Cleveland that night, for example, and their testimony is all I have got to go on in believing that he was in Cleveland, I might find that I no longer know whether he was there or not until I have some reason to rule out my suspicion. If their testimony were all invented I would not know that the man was in Cleveland. But strictly speaking his being in Cleveland is not incompatible with their making up a story saying he was. They might have invented a story to protect him, whereas in fact, unknown to them, he was there all the time. Such a complicated plot is not necessary to bring out the point; Moore's Duke of Devonshire is enough. From the fact that he was dreaming that he was speaking in the House of Lords it did not follow that he was not speaking in the House of Lords. In fact he was. The possibility of dreaming – which was actual in that case – did not imply the falsity of what was believed. A possible deficiency in the basis of my belief can interfere with my knowledge without itself rendering false the very thing I believe. A hallucinogenic drug might cause me to see my bed covered with a huge pile of leaves, for example.⁵ Having taken that drug, I will know the

actual state of my bed only if I know that what I see is not just the effect of the drug; I must be able to rule out the possibility that I am hallucinating the bed and the leaves. But however improbable it might be that my bed is actually covered with leaves, its not being covered with leaves does not follow from the fact that I am hallucinating that it is. What I am hallucinating could nevertheless be (unknown to me) true. But a goldfinch simply could not be a canary. So although there are two different ways in which a certain possibility can threaten my knowledge, it remains true that there are always certain possibilities which must be known not to obtain if I am to know what I claim to know.

I think these are just familiar facts about human knowledge, something we all recognize and abide by in our thought and talk about knowing things. We know what would be a valid challenge to a claim to know something, and we can recognize the relevance and force of objections made to our claims to know. The question before us is to what extent Descartes's investigation of his knowledge that he is sitting by the fire with a piece of paper in his hand follows these recognized everyday procedures for assessing claims to know. If it does follow them faithfully, and yet leads to the conclusion that he cannot know where he is or what is happening around him, we seem forced to accept his negative conclusion about knowledge in general just as we are forced to accept the conclusion that I do not know it is a goldfinch or do not know the witness was in Cleveland because I cannot rule out the possibilities which must be ruled out if I am to know such things. Is Descartes's introduction of the possibility that he might be dreaming just like the introduction of the possibility that it might be a canary in the garden or that the alibi might be contrived or that it might be a hallucination of my bed covered with leaves?

Those possibilities were all such that if they obtained I did not know what I claimed to know, and they had to be known not to obtain in order for the original knowledge-claim to be true. Does Descartes's dream-possibility fulfil both of those conditions? I have already said that it seems undeniable that it fulfils the first. If he *were* dreaming Descartes would not know what he claims to know. Someone who is dreaming does not thereby know anything about the world around him even if the world around him happens to be just the way he dreams or believes it to be. So his dreaming *is* incompatible with his knowing. But does it fulfil

the second condition? Is it a possibility which must be known not to obtain if Descartes is to know that he is sitting by the fire with a piece of paper in his hand? I think it is difficult simply to deny that it is. The evident force of Descartes's reasoning when we first encounter it is enough to show that it certainly strikes us as a relevant possibility, as something that he should know not to obtain if he is to know where he is and what is happening around him.

When that possibility strikes us as obviously relevant in Descartes's investigation we might come to think that it is because of a simple and obvious fact about knowledge. In the case of the goldfinch we immediately recognize that I must know that it is not a canary if I am to know it is a goldfinch. And it is very natural to think that that is simply because its being a canary is incompatible with its being a goldfinch. If it were a canary it would not be a goldfinch, and I would therefore be wrong in saying that it is; so if I am to know it is a goldfinch I must rule out the possibility that it is a canary. The idea is that the two conditions I distinguished in the previous paragraph are not really separate after all. As soon as we see that a certain possibility is incompatible with our knowing such-and-such, it is suggested, we immediately recognize that it is a possibility that must be known not to obtain if we are to know the such-and-such in question. We see that the dream-possibility satisfies that first condition in Descartes's case (if he were dreaming, he wouldn't know), and that is why, according to this suggestion, we immediately see that it is relevant and must be ruled out. Something we all recognize about knowledge is what is said to make that obvious to us.

But is the 'simple and obvious fact about knowledge' appealed to in this explanation really something that is true of human knowledge even in the most ordinary circumstances? What exactly is the 'fact' in question supposed to be? I have described it so far, as applied to the case of the goldfinch, as the fact that if I know something p (it's a goldfinch) I must know the falsity of all those things incompatible with p (e.g., it's a canary). If there were one of those things that I did not know to be false, and it were in fact true, I would not know that p , since in that case something incompatible with p would be true and so p would not be true. But to say that I must know that all those things incompatible with p are false is the same as saying that I must know the truth of all those things that must be true if p is true. And it is extremely

implausible to say that that is a 'simple and obvious fact' we all recognize about human knowledge.

The difficulty is that there are no determinate limits to the number of things that follow from the things I already know. But it cannot be said that I now know all those indeterminately many things, although they all must be true if the things that I already know are true. Even granting that I now know a great deal about a lot of different things, my knowledge obviously does not extend to everything that follows from what I now know. If it did, mathematics, to take only one example, would be a great deal easier than it is – or else impossibly difficult. In knowing the truth of the simple axioms of number theory, for example, I would thereby know the truth of everything that follows from them; every theorem of number theory would already be known. Or, taking the pessimistic side, since obviously no one does know all the theorems of number theory, it would follow that no one even knows that those simple axioms are true.

It is absurd to say that we enjoy or require such virtual omniscience, so it is more plausible to hold that the 'simple and obvious fact' we all recognize about knowledge is the weaker requirement that we must know the falsity of all those things that we *know* to be incompatible with the things we know. I know that a bird's being a canary is incompatible with its being a goldfinch; that is not some far-flung, unknown consequence of its being a goldfinch, but something that anyone would know who knew anything about goldfinches at all. And the idea is that that is why I must know that it is not a canary if I am to know that it is a goldfinch. Perhaps, in order to know something, p , I do not need to know the falsity of all those things that are incompatible with p , but it can seem that at least I must know the falsity of all those things that I *know* to be incompatible with p . Since I claim to know that the bird is a goldfinch, and I know that its being a goldfinch implies that it is not a canary, I must for that reason know that it is not a canary if my original claim is true. In claiming to know it is a goldfinch I was, so to speak, committing myself to knowing that it is not a canary, and I must honour my commitments.

This requirement as it stands, even if it does explain why I must know that the bird is not a canary, does not account for the relevance of the other sorts of possibilities I have mentioned. The reason in the goldfinch case was said to be that

I know that its being a canary is incompatible with its being a goldfinch. But that will not explain why I must rule out the possibility that the witnesses have invented a story about the man's being in Cleveland, or the possibility that I am hallucinating my bed covered with a pile of leaves. Nor will it explain why Descartes must rule out the possibility that he is dreaming. What I claimed to know in the first case is that the man was in Cleveland that night. But, as we saw earlier, it is not a consequence of his being in Cleveland that no one will invent a story to the effect that he was in Cleveland; they might mistakenly believe he was not there and then tell what they think is a lie. Nor is it a consequence of my bed's being covered with leaves that I am not hallucinating that it is. But we recognize that in order to know in those cases I nevertheless had to rule out those possibilities. Similarly, as the Duke of Devonshire reminds us, it is not a consequence of Descartes's sitting by the fire with a piece of paper in his hand that he is not dreaming that he is. So if it is obvious to us that Descartes must know that he is not dreaming if he is to know that he is sitting by the fire, it cannot be simply because the possibility in question is known to be incompatible with what he claims to know. It is not.

If there is some 'simple and obvious fact about knowledge' that we recognize and rely on in responding to Descartes's reasoning it must therefore be more complicated than what has been suggested so far. Reflecting even on the uncontroversial everyday examples alone can easily lead us to suppose that it is something like this: if somebody knows something, p , he must know the falsity of all those things incompatible with his knowing that p (or perhaps all those things he knows to be incompatible with his knowing that p). I will not speculate further on the qualifications or emendations needed to make the principle less implausible. The question now is whether it is our adherence to any such principle or requirement that is responsible for our recognition that the possibility that the bird is a canary or the possibility that the witnesses made up a story must be known not to obtain if I am to know the things I said I knew in those cases. What exactly are the procedures or standards we follow in the most ordinary, humdrum cases of putative knowledge? Reflection on the source of Descartes's sceptical reasoning has led to difficulties in describing and therefore in understanding even the most familiar procedures we follow in everyday life. That is one

of the rewards of a study of philosophical scepticism.

The main difficulty in understanding our ordinary procedures is that no principle like those I have mentioned could possibly describe the way we proceed in everyday life. Or, to put it less dogmatically, if our adherence to some such requirement were responsible for our reactions in those ordinary cases, Descartes would be perfectly correct, and philosophical scepticism about the external world would be true. Nobody would know anything about the world around us. If, in order to know something, we must rule out a possibility which is known to be incompatible with our knowing it, Descartes is perfectly right to insist that he must know that he is not dreaming if he is to know that he is sitting by the fire with a piece of paper in his hand. He knows his dreaming is incompatible with his knowing. I have already argued that if he is right in insisting that that condition must be fulfilled for knowledge of the world around us he is also right in concluding that it can never be fulfilled; fulfilling it would require knowledge which itself would be possible only if the condition were fulfilled. So both steps of Descartes's reasoning would be valid and his conclusion would be true.

That conclusion can be avoided, it seems to me, only if we can find some way to avoid the requirement that we must know we are not dreaming if we are to know anything about the world around us. But that requirement cannot be avoided if it is nothing more than an instance of a general procedure we recognize and insist on in making and assessing knowledge-claims in everyday and scientific life. We have no notion of knowledge other than what is embodied in those procedures and practices. So if that requirement is a 'fact' of our ordinary conception of knowledge we will have to accept the conclusion that no one knows anything about the world around us.

I now want to say a few more words about the position we would all be in if Descartes's conclusion as he understands it were correct. I described him earlier as having lost the whole world, as knowing at most what he is experiencing or how things appear to him, but knowing nothing about how things really are in the world around him. To show how anyone in that position could come to know anything about the world around him is what I am calling the problem of our knowledge of the external world, and it is worth dwelling for a

moment on just how difficult a problem that turns out to be if it has been properly raised.

If we are in the predicament Descartes finds himself in at the end of his *First Meditation* we cannot tell by means of the senses whether we are dreaming or not; all the sensory experiences we are having are compatible with our merely dreaming of a world around us while that world is in fact very different from the way we take it to be. Our knowledge is in that way confined to our sensory experiences. There seems to be no way of going beyond them to know that the world around us really is this way rather than that. Of course we might have very strongly held beliefs about the way things are. We might even be unable to get rid of the conviction that we are sitting by the fire holding a piece of paper, for example. But if we acknowledge that our sensory experiences are all we ever have to go on in gaining knowledge about the world, and we acknowledge, as we must, that given our experiences as they are we could nevertheless be simply dreaming of sitting by the fire, we must concede that we do not know that we are sitting by the fire. Of course, we are in no position to claim the opposite either. We cannot conclude that we are not sitting by the fire; we simply cannot tell which is the case. Our sensory experience gives us no basis for believing one thing about the world around us rather than its opposite, but our sensory experience is all we have got to go on. So whatever unshakeable conviction we might nevertheless retain, that conviction cannot be knowledge. Even if we are in fact holding a piece of paper by the fire, so that what we are convinced of is in fact true, that true conviction is still not knowledge. The world around us, whatever it might be like, is in that way beyond our grasp. We can know nothing of how it is, no matter what convictions, beliefs, or opinions we continue, perhaps inevitably, to hold about it.

What *can* we know in such a predicament? We can perhaps know what sensory experiences we are having, or how things seem to us to be. At least that much of our knowledge will not be threatened by the kind of attack Descartes makes on our knowledge of the world beyond our experiences. What we can know turns out to be a great deal less than we thought we knew before engaging in that assessment of our knowledge. Our position is much more restricted, much poorer, than we had originally supposed. We are confined at best to what Descartes calls 'ideas' of things around us, representations of things or states of affairs which,

for all we can know, might or might not have something corresponding to them in reality. We are in a sense imprisoned within those representations, at least with respect to our knowledge. Any attempt to go beyond them to try and tell whether the world really is as they represent it to be can yield only more representations, more deliverances of sense experience which themselves are compatible with reality's being very different from the way we take it to be on the basis of our sensory experiences. There is a gap, then, between the most that we can ever find out on the basis of our sensory experience and the way things really are. In knowing the one we do not thereby know the other.

This can seem to leave us in the position of finding a barrier between ourselves and the world around us. There would then be a veil of sensory experiences or sensory objects which we could not penetrate but which would be no reliable guide to the world beyond the veil. If we were in such a position, I think it is quite clear that we could not know what is going on beyond the veil. There would be no possibility of our getting reliable sensory information about the world beyond the veil; all such reports would simply be more representations, further ingredients of the ever-more-complicated veil. We could know nothing but the veil itself. We would be in the position of someone waking up to find himself locked in a room full of television sets and trying to find out what is going on in the world outside. For all he can know, whatever is producing the patterns he can see on the screens in front of him might be something other than well-functioning cameras directed on to the passing show outside the room. The victim might switch on more of the sets in the room to try to get more information, and he might find that some of the sets show events exactly similar or coherently related to those already visible on the screens he can see. But all those pictures will be no help to him without some independent information, some knowledge which does not come to him from the pictures themselves, about how the pictures he does see before him are connected with what is going on outside the room. The problem of the external world is the problem of finding out, or knowing how we could find out, about the world around us if we were in that sort of predicament. It is perhaps enough simply to put the problem this way to convince us that it can never be given a satisfactory solution.

But putting the problem this way, or only this way, has its drawbacks. For one thing, it encourages a facile dismissive response; not a solution to the problem as posed, but a rejection of it. I do not mean that we should not find a way to reject the problem – I think that is our only hope – but this particular response, I believe, is wrong, or at the very least premature. It is derived almost entirely from the perhaps overly dramatic description of the predicament I have just given.

I have described Descartes's sceptical conclusion as implying that we are permanently sealed off from a world we can never reach. We are restricted to the passing show on the veil of perception, with no possibility of extending our knowledge to the world beyond. We are confined to appearances we can never know to match or to deviate from the imperceptible reality that is forever denied us. This way of putting it naturally encourages us to minimize the seriousness of the predicament, to try to settle for what is undeniably available to us, or perhaps even to argue that nothing that concerns us or makes human life worthwhile has been left out.

If an imperceptible 'reality', as it is called on this picture, is forever inaccessible to us, what concern can it be of ours? How can something we can have no contact with, something from which we are permanently sealed off, even make sense to us at all? Why should we be distressed by an alleged limitation of our knowledge if it is not even possible for the 'limitation' to be overcome? If it makes no sense to aspire to anything beyond what is possible for us, it will seem that we should give no further thought to this allegedly imperceptible 'reality'. Our sensory experiences, past, present, and future, will then be thought to be all we are or should be concerned with, and the idea of a 'reality' lying beyond them necessarily out of our reach will seem like nothing more than a philosopher's invention. What a sceptical philosopher would be denying us would then be nothing we could have ordinary commerce with or interest in anyway. Nothing distressing about our ordinary position in the familiar world would have been revealed by a philosopher who simply invents or constructs something he calls 'reality' or 'the external world' and then demonstrates that we can have no access to it. That would show nothing wrong with the everyday sensory knowledge we seek and think we find in ordinary life and in scientific laboratories, nor would it show that our relation to the ordinary reality that concerns

us is different from what we originally thought it to be.

I think this reaction to the picture of our being somehow imprisoned behind the veil of our own sensory experiences is very natural and immediately appealing. It is natural and perhaps always advisable for a prisoner to try to make the best of the restricted life behind bars. But however much more bearable it makes the prospect of life-imprisonment, it should not lead him to deny the greater desirability, let alone the existence, of life outside. In so far as the comfort of this response to philosophical scepticism depends on such a denial it is at the very least premature and is probably based on misunderstanding. It depends on a particular diagnosis or account of how and why the philosophical argument succeeds in reaching its conclusion. The idea is that the 'conclusion' is reached only by contrivance. The inaccessible 'reality' denied to us is said to be simply an artefact of the philosopher's investigation and not something that otherwise should concern us. That is partly a claim about how the philosophical investigation of knowledge works; as such, it needs to be explained and argued for. We can draw no consolation from it until we have some reason to think it might be an accurate account of what the philosopher does. So far we have no such reason. On the contrary; so far we have every reason to think that Descartes has revealed the impossibility of the very knowledge of the world that we are most interested in and which we began by thinking we possess or can easily acquire. In any case, that would be the only conclusion to draw if Descartes's investigation does indeed parallel the ordinary kinds of assessments we make of our knowledge in everyday life.

We saw that I can ask what I really know about the common cold, or whether I really know that the witness was in Cleveland on the night in question, and that I can go on to discover that I do not really know what I thought I knew. In such ordinary cases there is no suggestion that what I have discovered is that I lack some special, esoteric thing called 'real knowledge', or that I lack knowledge of some exotic, hitherto-unheard-of domain called 'reality'. If I ask what I know about the common cold, and I come to realize that I do not really know whether it can be caused by sitting in a draught or not, the kind of knowledge I discover I lack is precisely what I was asking about or taking it for granted I had at the outset. I do not conclude with a shrug that it no longer matters because what I now find I lack is only knowledge about a special

domain called 'reality' that was somehow invented only to serve as the inaccessible realm of something called 'real knowledge'. I simply conclude that I don't really know whether colds are caused by sitting in draughts or not. If I say in a jury-room on Monday that we can eliminate the suspect because we know he was in Cleveland that night, and I then discover by reflection on Tuesday that I don't really know he was in Cleveland that night, what I am denying I have on Tuesday is the very thing I said on Monday that I had.

There is no suggestion in these and countless similar everyday cases that somehow in the course of our reflections on whether and how we know something we are inevitably led to change or elevate our conception of knowledge into something else called 'real knowledge' which we showed no signs of being interested in at the beginning. Nor is it plausible to suggest that our ordinary assessments of knowledge somehow lead us to postulate a 'reality' that is simply an artefact of our inquiries about our knowledge. When we ask whether we really know something we are simply asking whether we know that thing. The 'really' signifies that we have had second thoughts on the matter, or that we are subjecting it to more careful scrutiny, or that knowledge is being contrasted with something else, but not that we believe in something called 'real knowledge' which is different from or more elevated than the ordinary knowledge we are interested in. Knowing something differs from merely believing it or assuming it or taking it for granted or simply being under the impression that it is true, and so forth, so asking whether we really know something is asking whether we know it as opposed to, for example, merely believing it or assuming it or taking it for granted or simply being under the impression that it is true.

If that is true of our ordinary assessments of knowledge, and if Descartes's investigation of his knowledge that he is sitting by the fire with a piece of paper in his hand is just like those ordinary cases, his discovery that he doesn't know in the case he considers will have the same significance as it has in those ordinary cases. And if that example is indeed representative of our knowledge of the world around us, the kind of knowledge we are shown to lack will be the very kind of knowledge we originally thought we had of things like our sitting by the fire holding a piece of paper. Without a demonstration that Descartes's philosophical investigation differs from our ordinary assessments in some way that prevents its negative

conclusion from having the kind of significance similar conclusions are rightly taken to have in everyday life, we can derive no consolation from the ungrounded idea that the reality from which he shows our knowledge is excluded does not or should not concern us anyway. It is the investigation of his everyday knowledge, and not merely the fanciful picture of a veil of perception, that generates Descartes's negative verdict.

But even if we did try to console ourselves with the thought that we can settle for what we *can* know on Descartes's account, how much consolation could it give us? The position Descartes's argument says we are in is much worse than what is contemplated in the optimistic response of merely shrugging off any concern with an imperceptible 'reality'.

For one thing, we would not in fact be left with what we have always taken to be the familiar objects of our everyday experience – tables and chairs, trees and flowers, bread and wine. If Descartes is right, we know nothing of such things. What we perceive and are in direct sensory contact with is never a physical object or state of affairs, but only a representation – something that could be just the way it is even if there were no objects at all of the sort it represents. So if we were to settle for the realm of things we could have knowledge about even if Descartes's conclusion were correct, we would not be settling for the comfortable world with which we began. We would have lost all of that, at least as something we can know anything about, and we would be restricted to facts about how things seem to us at the moment rather than how they are.

It might still be felt that after all nothing is certain in this changing world, so we should not aspire to firm truths about how things are. As long as we know that all or most of us agree about how things seem to us, or have seemed to us up till now, we might feel we have enough to give our social, cultural, and intellectual life as much stability as we can reasonably expect or need. But again this reaction *does not* really acknowledge the poverty or restrictedness of the position Descartes's sceptical conclusion would leave each of us in. Strictly speaking, there is no community of acting, experiencing and thinking persons I can know anything about if Descartes is correct. Other people, as I understand them, are not simply sensory experiences of mine; they too, if they exist, will therefore inhabit the unreachable world beyond my sensory experiences, along with the tables and

chairs and other things about which I can know nothing. So at least with respect to what I can know I could not console myself with thoughts of a like-minded community of perceivers all working together and cheerfully making do with what a communal veil of perception provides. I would have no more reason to believe that there are any other people than I have to believe that I am now sitting in a chair writing. The representations or sensory experiences to which Descartes's conclusion would restrict my knowledge could be no other than my own sensory experiences; there could be no communal knowledge even of the veil of perception itself. If my own sensory experiences do not make it possible for me to know things about the world around me they do not make it possible for me to know even whether

there are any other sensory experiences or any other perceiving beings at all.

The consequences of accepting Descartes's conclusion as it is meant to be understood are truly disastrous. There is no easy way of accommodating oneself to its profound negative implications. But perhaps by now we have come far enough to feel that the whole idea is simply absurd, that ultimately it is not even intelligible, and that there can be no question of 'accepting' Descartes's conclusion at all. I have no wish to discourage such a reaction. I would only insist that the alleged absurdity or unintelligibility must be identified and made out. I think that is the only way we can hope to learn whatever there is to be learned from Descartes's investigation.

Notes

- 1 It has been argued that the problem in the completely general form in which I discuss it here is new in Descartes, and that nothing exactly similar appears in philosophy before that time. See M. F. Burnyeat, 'Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed', *The Philosophical Review* (1982).
- 2 See the beginning of the first of his *Meditations on First Philosophy* in *The Philosophical Works of Descartes*, edited and translated by E. S. Haldane and G.

R. T. Ross (2 vols, New York, 1955), vol. I, p. 145. (Hereafter cited as HR.)

- 3 See his *Discourse on the Method of Rightly Conducting Reason and Seeking Truth in the Sciences* in HR, pp. 81ff.
- 4 See G. E. Moore, 'Certainty', this vol., ch. 4.
- 5 A memorable example H. H. Price gave in a lecture in 1962. It is my impression that Price was reporting on an actual hallucination of his.

Proof of an External World

G. E. Moore

It seems to me that, so far from its being true, as Kant declares to be his opinion, that there is only one possible proof of the existence of things outside of us, namely the one which he has given, I can now give a large number of different proofs, each of which is a perfectly rigorous proof; and that at many other times I have been in a position to give many others. I can prove now, for instance, that two human hands exist. How? By holding up my two hands, and saying, as I make a certain gesture with the right hand, 'Here is one hand', and adding, as I make a certain gesture with the left, 'and here is another'. And if, by doing this, I have proved *ipso facto* the existence of external things, you will all see that I can also do it now in numbers of other ways: there is no need to multiply examples.

But did I prove just now that two human hands were then in existence? I do want to insist that I did; that the proof which I gave was a perfectly rigorous one; and that it is perhaps impossible to give a better or more rigorous proof of anything whatever. Of course, it would not have been a proof unless three conditions were satisfied; namely (1) unless the premiss which I adduced as proof of the conclusion was different from the conclusion I adduced it to prove; (2) unless the premiss which I adduced was something which I *knew* to be the case, and not merely something which I believed but which was by no means certain, or something which, though in fact true, I did not know to be so; and (3) unless the conclusion did really follow from the premiss. But all

these three conditions were in fact satisfied by my proof. (1) The premiss which I adduced in proof was quite certainly different from the conclusion, for the conclusion was merely 'Two human hands exist at this moment'; but the premiss was something far more specific than this – something which I expressed by showing you my hands, making certain gestures, and saying the words 'Here is one hand, and here is another'. It is quite obvious that the two were different, because it is quite obvious that the conclusion might have been true, even if the premiss had been false. In asserting the premiss I was asserting much more than I was asserting in asserting the conclusion. (2) I certainly did at the moment *know* that which I expressed by the combination of certain gestures with saying the words 'Here is one hand and here is another'. I *knew* that there was one hand in the place indicated by combining a certain gesture with my first utterance of 'here' and that there was another in the different place indicated by combining a certain gesture with my second utterance of 'here'. How absurd it would be to suggest that I did not know it, but only believed it, and that perhaps it was not the case! You might as well suggest that I do not know that I am now standing up and talking – that perhaps after all I'm not, and that it's not quite certain that I am! And finally (3) it is quite certain that the conclusion did follow from the premiss. This is as certain as it is that if there is one hand here and another here *now*, then it follows that there are two hands in existence *now*.

My proof, then, of the existence of things outside of us did satisfy three of the conditions necessary for a rigorous proof. Are there any other

From G. E. Moore, *Philosophical Papers* (New York: Collier Books, 1962), pp. 144–6.

conditions necessary for a rigorous proof, such that perhaps it did not satisfy one of them? Perhaps there may be; I do not know; but I do want to emphasise that, so far as I can see, we all of us do constantly take proofs of this sort as absolutely conclusive proofs of certain conclusions – as finally settling certain questions, as to which we were previously in doubt. Suppose, for instance, it were a question whether there were as many as three misprints on a certain page in a certain book. A says there are, B is inclined to doubt it. How could A prove that he is right? Surely he *could* prove it by taking the book, turning to the page, and pointing to three separate places on it, saying ‘There’s one misprint here, another here, and another here’: surely that is a method by which it *might* be proved! Of course, A would not have proved, by doing this, that there were at least three misprints on the page in question, unless it were certain that there was a misprint in each of the places to which he pointed. But to say that he *might* prove it in this way, is to say that it *might* be certain that there was. And if such a thing as that could ever be certain, then assuredly it was certain just now that there was one hand in one of the two places I indicated and another in the other.

I did, then, just now, give a proof that there were *then* external objects; and obviously, if I did, I could *then* have given many other proofs of the same sort that there were external objects *then*, and could now give many proofs of the same sort that there are external objects *now*.

But, if what I am asked to do is to prove that external objects have existed *in the past*, then I can give many different proofs of this also, but proofs which are in important respects of a different *sort* from those just given. And I want to emphasise that, when Kant says it is a scandal not to be able to give a proof of the existence of external objects, a proof of their existence in the past would certainly *help* to remove the scandal of which he is speaking. He says that, if it occurs to anyone to question their existence, we ought to be able to confront him with a satisfactory proof. But by a person who questions their existence, he certainly means not merely a person who questions whether any exist at the moment of speaking, but a person who questions whether any have *ever* existed; and a proof that some have existed in the past would certainly therefore be relevant to *part* of what such a person is questioning. How then can I prove that there have been external objects in the

past? Here is one proof. I can say: ‘I held up two hands above this desk not very long ago; therefore two hands existed not very long ago; therefore at least two external objects have existed at some time in the past, QED’. This is a perfectly good proof, provided I *know* what is asserted in the premiss. But I *do* know that I held up two hands above this desk not very long ago. As a matter of fact, in this case you all know it too. There’s no doubt whatever that I did. Therefore I have given a perfectly conclusive proof that external objects have existed in the past; and you will all see at once that, if this is a conclusive proof, I could have given many others of the same sort, and could now give many others. But it is also quite obvious that this sort of proof differs in important respects from the sort of proof I gave just now that there were two hands existing *then*.

I have, then, given two conclusive proofs of the existence of external objects. The first was a proof that two human hands existed at the time when I gave the proof; the second was a proof that two human hands had existed at a time previous to that at which I gave the proof. These proofs were of a different sort in important respects. And I pointed out that I could have given, then, many other conclusive proofs of both sorts. It is also obvious that I could give many others of both sorts now. So that, if these are the sort of proof that is wanted, nothing is easier than to prove the existence of external objects.

But now I am perfectly well aware that, in spite of all that I have said, many philosophers will still feel that I have not given any satisfactory proof of the point in question. And I want briefly, in conclusion, to say something as to why this dissatisfaction with my proofs should be felt.

One reason why, is, I think, this. Some people understand ‘proof of an external world’ as including a proof of things which I haven’t attempted to prove and haven’t proved. It is not quite easy to say *what* it is that they want proved – *what* it is that is such that unless they got a proof of it, they would not say that they had a proof of the existence of external things; but I can make an approach to explaining what they want by saying that if I had proved the propositions which I used as *premises* in my two proofs, then they would perhaps admit that I had proved the existence of external things, but, in the absence of such a proof (which, of course, I have neither given nor attempted to give), they will say that I have not given what they mean by a proof of the existence of

external things. In other words, they want a proof of what I assert *now* when I hold up my hands and say 'Here's one hand and here's another'; and, in the other case, they want a proof of what I assert *now* when I say 'I did hold up two hands above this desk just now'. Of course, what they really want is not merely a proof of these two propositions, but something like a general statement as to how *any* propositions of this sort may be proved. This, of course, I haven't given; and I do not believe it can be given: if this is what is meant by proof of the existence of external things, I do not believe that any proof of the existence of external things is possible. Of course, in some cases what might be called a proof of propositions which seem like these can be got. If one of you suspected that one of my hands was artificial he might be said to get a proof of my proposition 'Here's one hand, and here's another', by coming up and examining the suspected hand close up, perhaps touching and pressing it, and so establishing that it really was a human hand. But I do not believe that any proof is possible in nearly all cases. How am I to prove now that 'Here's one hand, and here's another'? I do not believe I can do it. In order to do it, I should need to prove for one thing, as Descartes pointed out, that I am not now dreaming. But how can I prove that I am not? I have, no doubt, conclusive reasons for asserting that I am not now dreaming; I have conclusive evidence that I am awake: but that is a very different thing from being able to prove it. I could not tell you what all my evidence is; and I should require to do this at least, in order to give you a proof.

But another reason why some people would feel dissatisfied with my proofs is, I think, not merely that they want a proof of something which I haven't proved, but that they think that, if I cannot give such extra proofs, then the proofs that I have given are not conclusive proofs at all. And this, I think, is a definite mistake. They would say: 'If you cannot prove your premiss that here is one hand and here is another, then you do not know it. But you yourself have admitted that, if you did not know it, then your proof was not conclusive. Therefore your proof was not, as you say it was, a conclusive proof.' This view that, if I cannot prove such things as these, I do not know them, is, I think, the view that Kant was expressing in the sentence which I quoted at the beginning of this lecture, when he implies that so long as we have no proof of the existence of external things, their existence must be accepted merely on *faith*. He means to say, I think, that if I cannot prove that there is a hand here, I must accept it merely as a matter of faith – I cannot know it. Such a view, though it has been very common among philosophers, can, I think, be shown to be wrong – though shown only by the use of premisses which are not known to be true, unless we do know of the existence of external things. I can know things, which I cannot prove; and among things which I certainly did know, even if (as I think) I could not prove them, were the premisses of my two proofs. I should say, therefore, that those, if any, who are dissatisfied with these proofs merely on the ground that I did not know their premisses, have no good reason for their dissatisfaction.

Four Forms of Scepticism

G. E. Moore

We pass next to the argument: "Descartes's malicious demon is a logical possibility." This is obviously quite different from both the two preceding. Russell does not say that any percepts *are* produced by Descartes's malicious demon; nor does he mean that it is practically or theoretically possible for Descartes's malicious demon to produce in me percepts like this, in the sense in which it is (perhaps) practically possible that a conjurer should, and theoretically possible that a physiologist should by stimulating the optic nerve. He only says it is a *logical possibility*. But what exactly does this mean? It is, I think, an argument which introduces quite new considerations, of which I have said nothing so far, and which lead us to the root of the difference between Russell and me. I take it that Russell is here asserting that it is *logically possible* that this particular percept of mine, which I think I know to be associated with a percept belonging to someone else, was in fact produced in me by a malicious demon when there was no such associated percept. and that, therefore, I cannot know for certain what I think I know. It is, of course, being assumed that, *if* it was produced by a malicious demon, then it follows that it is not associated with a percept belonging to someone else, in the way in which I think I know it is: that is how the phrase "was produced by a malicious demon" is being used. The questions we have to consider are, then, simply these three: What is meant by saying that it is *logically possible* that this percept was produced by a malicious demon? Is it *true* that this is logically

possible? And: If it is true, does it follow that I don't know for certain that it was *not* produced by a malicious demon?

Now there are three different things which might be meant by saying that this proposition is logically possible. The first is that it is not a self-contradictory proposition. This I readily grant. But from the mere fact that it is not self-contradictory, it certainly does not follow that I don't know for certain that it is false. This Russell grants. He holds that I do know for certain to be false, propositions about my percepts which are not self-contradictory. He holds, for instance, that I do know for certain that there is a white visual percept now; and yet the proposition that there isn't is certainly not self-contradictory.

He must, therefore, in his argument, be using "logically possible" in some other sense. And one sense in which it might naturally be used is this: Not logically incompatible with anything that I know. If, however, he were using it in this sense, he would be simply begging the question. For the very thing I am claiming to know is that this percept was *not* produced by a malicious demon and of course the proposition that it was produced by a malicious demon *is* incompatible with the proposition that it was *not*.

There remains one sense, which is, I think, the sense in which he is actually using it. Namely he is saying: The proposition "This percept was produced by a malicious demon" is *not* logically incompatible with anything you know *immediately*. And if this is what he means, I own that I think Russell is right. This is a matter about which I suppose many philosophers would disagree with us. There are people who suppose that I *do* know

immediately, in certain cases, such things as: That person is conscious; at least, they use this language, though whether they mean exactly what I am here meaning by "know immediately" may be doubted. I can, however, not help agreeing with Russell that I never do know *immediately* that that person is conscious, nor anything else that is *logically incompatible* with "This percept was produced by a malicious demon." Where, therefore, I differ from him is in supposing that I do know for certain things which I do not know immediately and which also do *not* follow logically from anything which I do know immediately.

This seems to me to be the fundamental question at issue in considering my classes (3) and (4) and what distinguishes them from cases (1) and (2). I think I do know *immediately* things about myself and such things as "There was a sound like 'Russell' a little while ago" – that is, I think that memory is *immediate* knowledge and that much of my knowledge about myself is immediate. But I cannot help agreeing with Russell that I never know immediately such a thing as "That person is conscious" or "This is a pencil," and that also the truth of such propositions never follows logically from anything which I do know immediately, and yet I think that I do know such things for certain. Has he any argument for his view that if their falsehood is *logically possible* (i.e. if I do not know *immediately* anything logically incompatible with their falsehood) then I do *not* know them for certain? This is a thing which he certainly constantly assumes; but I cannot find that he anywhere gives any distinct arguments for it.

So far as I can gather, his reasons for holding it are the two assumptions which he expresses when he says: "If (I am to reject the view that my life is one long dream) I must do so on the basis of an analogical or inductive argument, which cannot give complete certainty."¹ That is to say he assumes: (1) My belief or knowledge that this is a pencil is, *if* I do not know it immediately, and if also the proposition does not follow logically from

anything that I know immediately, in some sense "based on" an analogical or inductive argument; and (2) What is "based on" an analogical or inductive argument is never certain knowledge, but only more or less probable belief. And with regard to these assumptions, it seems to me that the first must be true in some sense or other, though it seems to me terribly difficult to say exactly what the sense is. What I am inclined to dispute, therefore, is the second: I am inclined to think that what is "based on" an analogical or inductive argument, in the sense in which my knowledge or belief that this is a pencil is so, may nevertheless be certain knowledge and *not* merely more or less probable belief.

What I want, however, finally to emphasize is this: Russell's view that I do not know for certain that this is a pencil or that you are conscious rests, if I am right, on no less than four distinct assumptions: (1) That I don't know these things immediately; (2) That they don't follow logically from any thing or things that I do know immediately; (3) That, *if* (1) and (2) are true, my belief in or knowledge of them must be "based on an analogical or inductive argument"; and (4) That what is so based cannot be *certain knowledge*. And what I can't help asking myself is this: Is it, in fact, as certain that all these four assumptions are true, as that I *do* know that this is a pencil and that you are conscious? I cannot help answering: It seems to me *more* certain that I *do* know that this is a pencil and that you are conscious, than that any single one of these four assumptions is true, let alone all four. That is to say, though, as I have said, I agree with Russell that (1), (2) and (3) *are* true; yet of no one even of these three do I feel *as* certain as that I do know for certain that this is a pencil. Nay more: I do not think it is *rational* to be as certain of any one of these four propositions, as of the proposition that I do know that this is a pencil. And how on earth is it to be decided which of the two things it is *rational* to be most certain of?

Notes

1 Bertrand Russell, *An Outline of Philosophy* (Allen & Unwin: London, 1927), p. 218.

Certainty

G. E. Moore

Suppose I say: "I know for certain that I am standing up; it is absolutely certain that I am; there is not the smallest chance that I am not." Many philosophers would say: "You are wrong: you do not know that you are standing up; it is *not* absolutely certain that you are; there is *some* chance, though perhaps only a very small one, that you are not." And one argument which has been used as an argument in favour of saying this, is an argument in the course of which the philosopher who used it would assert: "You do not know for certain that you are not dreaming; it is not absolutely certain that you are not; there is *some* chance, though perhaps only a very small one, that you are." And from this, that I do not know for certain that I am not dreaming, it is supposed to follow that I do not know for certain that I am standing up. It is argued: If it is not certain that you are not dreaming, then it is not certain that you are standing up. And that if I don't know that I'm not dreaming, I also don't know that I'm not sitting down, I don't feel at all inclined to dispute. From the hypothesis that I am dreaming, it would, I think, certainly follow that I don't *know* that I am standing up; though I have never seen the matter argued, and though it is not at all clear to me how it is to be proved that it would follow. But, on the other hand, from the hypothesis that I am dreaming, it certainly would not follow that I am *not* standing up; for it is certainly logically possible that a man should be fast asleep and dreaming, while he is standing up and not lying down. It is therefore logically possible that I should both be

standing up and also at the same time dreaming that I am; just as the story, about a well-known Duke of Devonshire, that he once dreamt that he was speaking in the House of Lords and, when he woke up, found that he *was* speaking in the House of Lords, is certainly logically possible. And if, as is commonly assumed, when I am dreaming that I am standing up it may also be correct to say that I am *thinking* that I am standing up, then it follows that the hypothesis that I am now dreaming is quite consistent with the hypothesis that I am both thinking that I am standing up and also actually standing up. And hence, if as seems to me to be certainly the case and as this argument assumes, from the hypothesis that I am now dreaming it *would* follow that I don't know that I am standing up, there follows a point which is of great importance with regard to our use of the word "knowledge," and therefore also of the word "certainty" – a point which has been made quite conclusively more than once by Russell, namely that from the conjunction of the two facts that a man thinks that a given proposition *p* is true, and that *p* is in fact true, it does *not* follow that the man in question *knows* that *p* is true: in order that I may be justified in saying that I know that I am standing up, something more is required than the mere conjunction of the two facts that I both think I am and actually am – as Russell has expressed it, true belief is not identical with knowledge; and I think we may further add that even from the conjunction of the two facts that I feel certain that I am and that I actually am it would not follow that I know that I am, nor therefore that it is certain that I am. As regards the argument drawn from the

fact that a man who dreams that he is standing up and happens at the moment actually to be standing up will nevertheless not *know* that he is standing up, it should indeed be noted that from the fact that a man is dreaming that he is standing up, it certainly does not *follow* that he *thinks* he is standing up; since it does sometimes happen in a dream that we *think* that it is a dream, and a man who thought this certainly might, although he was dreaming that he was standing up, yet *think* that he was not, although he could not *know* that he was not. It is not therefore the case, as might be hastily assumed, that, if I dream that I am standing up at a time when I am in fact lying down, I am necessarily *deceived*: I should be deceived only if I thought I was standing when I wasn't; and I may dream that I am, without *thinking* that I am. It certainly does, however, often happen that we do dream that so-and-so is the case, without at the time thinking that we are only dreaming; and in such cases, I think we may perhaps be said to *think* that what we dream is the case *is* the case, and to be deceived if it is not the case; and therefore also, in such cases, if what we dream to be the case happens also to *be* the case, we may be said to be thinking truly that it is the case, although we certainly do not *know* that it is.

I agree, therefore, with that part of this argument which asserts that if I don't know now that I'm not dreaming, it follows that I don't *know* that I am standing up, even if I both actually am and think that I am. But this first part of the argument is a consideration which cuts both ways. For, if it is true, it follows that it is also true that if I *do* know that I am standing up, then I do know that I am not dreaming. I can therefore just as well argue: since I do know that I'm standing up, it follows that I do know that I'm not dreaming; as my opponent can argue: since you don't know that you're not dreaming, it follows that you don't know that you're standing up. The one argument is just as good as the other, unless my opponent can give better reasons for asserting that I don't know that I'm not dreaming, than I can give for asserting that I do know that I am standing up.

What reasons can be given for saying that I don't know for certain that I'm not at this moment dreaming?

I do not think that I have ever seen clearly stated any argument which is supposed to show this. But I am going to try to state, as clearly as I can, the premisses and the reasonings from them, which I think have led so many philosophers to suppose

that I really cannot now know for certain that I am not dreaming.

I said, you may remember, in talking of the seven assertions with which I opened this lecture, that I had "the evidence of my senses" for them, though I also said that I didn't think this was the only evidence I had for them, nor that this by itself was necessarily conclusive evidence. Now if I had *then* "the evidence of my senses" in favour of the proposition that I was standing up, I certainly have *now* the evidence of my senses in favour of the proposition that I *am* standing up, even though this may not be all the evidence that I have, and may not be conclusive. But have I, in fact, the evidence of my senses *at all* in favour of this proposition? One thing seems to me to be quite clear about our use of this phrase, namely, that, if a man at a given time is only dreaming that he is standing up, then it follows that he has *not* at that time the evidence of his senses in favour of that proposition: to say "Jones last night was *only* dreaming that he was standing up, and yet all the time he had the evidence of his senses that he was" is to say something self-contradictory. But those philosophers who say it is possible that I am now dreaming, certainly mean to say also that it is possible that I am *only dreaming* that I am standing up; and this view, we now see, entails that it is possible that I have *not* the evidence of my senses that I am. If, therefore, they are right, it follows that it is not certain even that I have the evidence of my senses that I am; it follows that it is not certain that I have the *evidence of my senses* for anything at all. If, therefore, I were to say now, that I certainly have the evidence of my senses in favour of the proposition that I am standing up, even if it's not certain that I am standing up, I should be begging the very question now at issue. For if it is not certain that I am not dreaming, it is not certain that I even have the evidence of my senses that I am standing up.

But, now, even if it is not certain that I have at this moment the evidence of my senses for anything at all, it is quite certain that I *either* have the evidence of my senses that I am standing up *or* have an experience which is *very like* having the evidence of my senses that I am standing up. *If* I am dreaming, this experience consists in having dream-images which are at least very like the sensations I should be having if I were awake and had the sensations, the having of which would constitute "having the evidence of my senses" that I am standing up. Let us use the expression "sensory

experience," in such a way that this experience which I certainly am having will be a "sensory experience," whether or not it merely consists in the having of dream-images. If we use the expression "sensory experience" in this way, we can say, I think, that, if it is not certain that I am not dreaming now, then it is not certain that *all* the sensory experiences I am now having are not mere dream-images.

What then are the premisses and the reasonings which would lead so many philosophers to think that all the sensory experiences I am having now *may* be mere dream-images – that I do not know for certain that they are not?

So far as I can see, one premiss which they would certainly use would be this: "Some at least of the sensory experiences which you are having now are similar in important respects to dream-images which actually have occurred in dreams." This seems a very harmless premiss, and I am quite willing to admit that it is true. But I think there is a very serious objection to the procedure of using it as a premiss in favour of the derived conclusion. For a philosopher who does use it as a premiss, is, I think, in fact *implying*, though he does not expressly say, that he himself knows it to be true. He is *implying* therefore that he himself knows that dreams have occurred. And, of course, I think he would be right. All the philosophers I have ever met or heard of certainly did know that dreams have occurred: we all know that dreams *have* occurred. But can he consistently combine this proposition that he knows that dreams have occurred, with his conclusion that he does not know that he is not dreaming? Can anybody possibly know that dreams have occurred, if, at the time, he does not himself know that he is not dreaming? If he *is* dreaming, it may be that he is only dreaming that dreams have occurred; and if he does not know that he is not dreaming, can he possibly know that he is *not* only dreaming that dreams have occurred? Can he possibly know therefore that dreams *have* occurred? I do not think that he can; and therefore I think that anyone who uses this premiss and also asserts the conclusion that nobody ever knows that he is not dreaming, is guilty of an inconsistency. By using this premiss he *implies* that he himself knows that dreams have occurred; while, if his conclusion is true, it follows that he himself does not know that he is not dreaming, and therefore does not know that he is not only dreaming that dreams have occurred.

However, I admit that the premiss is true. Let us now try to see by what sort of reasoning it might be thought that we could get from it to the conclusion.

I do not see how we can get forward in that direction at all, unless we first take the following huge step, unless we say, namely: since there have been dream-images similar in important respects to some of the sensory experiences I am now having, it is logically possible that there should be dream-images *exactly like all* the sensory experiences I am now having, and logically possible, therefore, that all the sensory experiences I am now having *are* mere dream-images. And it might be thought that the validity of this step could be supported to some extent by appeal to matters of fact, though only, of course, at the cost of the same sort of inconsistency which I have just pointed out. It might be said, for instance, that some people have had dream-images which were *exactly like* sensory experiences which they had when they were awake, and that therefore it must be logically possible to have a dream-image exactly like a sensory experience which is *not* a dream-image. And then it may be said: If it is logically possible for some dream-images to be exactly like sensory experiences which are not dream-images, surely it must be logically possible for *all* the dream-images occurring in a dream at a given time to be exactly like sensory experiences which are not dream-images, and logically possible also for all the sensory experiences which a man has at a given time when he is awake to be exactly like all the dream-images which he himself or another man had in a dream at another time.

Now I cannot see my way to deny that it is logically possible that all the sensory experiences I am having now should be mere dream-images. And if this is logically possible, and if further the sensory experiences I am having now were the only experiences I am having, I do not see how I could possibly know for certain that I am not dreaming.

But the conjunction of my memories of the immediate past with these sensory experiences *may* be sufficient to enable me to know that I am not dreaming. I say it *may* be. But what if our sceptical philosopher says: It is *not* sufficient; and offers as an argument to prove that it is not, this: It is logically possible *both* that you should be having all the sensory experiences you are having, and also that you should be remembering what you do remember, and *yet* should be dreaming. If this is

G. E. Moore

logically possible, then I don't see how to deny that I cannot possibly know for certain that I am not dreaming: I do not see that I possibly could. But can any reason be given for saying that it *is* logically possible? So far as I know nobody ever has, and I don't know how anybody ever could. And so long as this is not done my argument, "I know that I am standing up, and therefore I know that I am not dreaming," remains at least as good as his, "You don't know that you are not dreaming, and therefore don't know that you are standing up." And I don't think I've ever seen an argument expressly directed to show that it is not.

One final point should be made clear. It is certainly logically possible that I *should have* been dreaming now; I *might* have been dreaming now; and therefore the proposition that I *am* dreaming now is not self-contradictory. But what I am in doubt of is whether it is logically possible that I should *both* be having all the sensory experiences and the memories that I have and *yet* be dreaming. The conjunction of the proposition that I have these sense experiences and memories with the proposition that I am dreaming does seem to me to be very likely self-contradictory.

Skepticism, Naturalism and Transcendental Arguments

P. F. Strawson

1 Introductory Remarks

The term "naturalism" is elastic in its use. The fact that it has been applied to the work of philosophers having as little in common as Hume and Spinoza is enough to suggest that there is a distinction to be drawn between varieties of naturalism. In later chapters, I shall myself draw a distinction between two main varieties, within which there are subvarieties. Of the two main varieties, one might be called *strict* or *reductive* naturalism (or, perhaps, *hard* naturalism). The other might be called *catholic* or *liberal* naturalism (or, perhaps, *soft* naturalism). The words "catholic" and "liberal" I use here in their comprehensive, not in their specifically religious or political, senses; nothing I say will have any direct bearing on religion or the philosophy of religion or on politics or political philosophy.

Each of these two general varieties of naturalism will be seen by its critics as liable to lead its adherents into intellectual aberration. The exponent of some subvarieties of strict or reductive naturalism is liable to be accused of what is pejoratively known as scientism, and of denying evident truths and realities. The soft or catholic naturalist, on the other hand, is liable to be accused of fostering illusions or propagating *myths*. I do not want to suggest that a kind of intellectual cold war between the two is inevitable. There is, perhaps, a

possibility of compromise or *détente*, even of reconciliation. The soft or catholic naturalist, as his name suggests, will be the reader with proposals for peaceful coexistence.

My title seems to speak of varieties of skepticism as well as varieties of naturalism. An exponent of some subvariety of reductive naturalism in some particular area of debate may sometimes be seen, or represented, as a kind of skeptic in that area: say, a moral skeptic or a skeptic about the mental or about abstract entities or about what are called "intensions." I shall explore some of these areas later on; and it is only then that the distinction between hard and soft naturalism will come into play.

For the present, I shall not need any such distinction and I shall not make any such slightly deviant or extended applications of the notion of skepticism. To begin with, I shall refer only to some familiar and standard forms of philosophical skepticism. Strictly, skepticism is a matter of doubt rather than of denial. The skeptic is, strictly, not one who denies the validity of certain types of belief, but one who questions, if only initially and for methodological reasons, the adequacy of our grounds for holding them. He puts forward his doubts by way of a challenge — sometimes a challenge to himself — to show that the doubts are unjustified, that the beliefs put in question are justified. He may conclude, like Descartes, that the challenge can successfully be met; or, like Hume, that it cannot (though this view of Hume's was importantly qualified). Traditional targets of philosophic doubt include the existence of the external world, i.e. of physical objects or bodies; our knowledge of other minds; the justification of

induction; the reality of the past. Hume concerned himself most with the first and third of these – body and induction; and I shall refer mainly, though not only, to the first.

I shall begin by considering various different kinds of attempts to meet the challenge of traditional skepticism by argument; and also various replies to these attempts, designed to show that they are unsuccessful or that they miss the point. Then I shall consider a different kind of response to skepticism – a response which does not so much attempt to meet the challenge as to pass it by. And this is where I shall first introduce an undifferentiated notion of Naturalism. The hero of this part of the story is Hume: he appears in the double role of arch-skeptic and arch-naturalist. Other names which will figure in the story include those of Moore, Wittgenstein, Carnap and, among our own contemporaries, Professor Barry Stroud. This part of the story is the theme of the present chapter. It is an old story, so I shall begin by going over some familiar ground.

2 Traditional Skepticism

To begin, then, with G. E. Moore. It will be remembered that in his famous “A Defence of Common Sense”¹ Moore asserted that he, and very many other people as well, knew with certainty a number of propositions regarding which some philosophers had held that they were not, and could not be, known with certainty. These propositions included the proposition that the earth had existed for a great many years; that on it there had been, and were now, many bodies, or physical objects, of many different kinds; that these bodies included the bodies of human beings who, like Moore himself, had had, or were having, thoughts and feelings and experiences of many different kinds. If Moore was right in holding that such propositions are widely known, with certainty, to be true, then it seems to follow that certain theses of philosophical skepticism are false: e.g. the thesis that it cannot be known with certainty that material objects exist, and the thesis that no one can know with certainty of the existence of any minds other than his own or, to put it a little more bluntly, that no one can know with certainty that there are other people. Again, the first of these two skeptical theses is implicitly challenged, indeed denied, by Moore in yet another famous paper called “Proof of an External

World.”² He claimed, in delivering this paper, to prove that two human hands exist, hence that external things exist, by holding up first one hand, then another and saying, as he did so, “Here is one hand and here is another.” The proof was rigorous and conclusive, he claimed, since he knew for certain that the premise was true and it was certain that the conclusion followed from the premise.

It was hardly to be expected that Moore’s “Defence” or his “Proof” would be universally accepted as settling the questions to which they were addressed. Rather, it was felt by some philosophers that the point of philosophical skepticism about, say, the existence of external things, of the physical world, was somehow being missed. A recent expression of this feeling is given by Professor Barry Stroud in an article called “The Significance of Scepticism.”³ At its most general, the skeptical point concerning the external world seems to be that subjective experience could, logically, be just the way it is without its being the case that physical or material things actually existed. (Thus Berkeley, for example, embraced a different hypothesis – that of a benevolent deity as the cause of sense-experiences – and we can find in Descartes the suggestion, though not, of course, the endorsement, of another – that of a malignant demon; while the consistent phenomenalist questions the need for any external source of sense-experience at all.) So if Moore, in making the claims he made, was simply relying on his own experience being just the way it was, he was missing the skeptical point altogether; and if he was not, then, since he issues his knowledge-claims without any further argument, all he has done is simply to issue a dogmatic denial of the skeptical thesis. But simple dogmatism settles nothing in philosophy. Stroud, at the end of his article, suggests that we ought to try to find some way of *defusing* skepticism. He does not mean, some way of establishing or proving that we do know for certain what the skeptic denies we know for certain, for he does not appear to think that this is possible; but, rather, some way of *neutralizing* the skeptical question, rendering it philosophically *impotent*. These expressions are not very clear, but I doubt if Stroud intended them to be.

Stroud mentions one attempt to neutralize the skeptical question, an attempt which he finds unsatisfactory. The attempt is Carnap’s.⁴ Carnap distinguished two ways in which the words “There

are or exist external or physical things” might be taken. On one interpretation these words simply express a proposition which is an obvious truism, a trivial consequence of hosts of propositions, like Moore’s “Here are two hands,” which are ordinarily taken, and in a sense correctly taken, to be empirically verified, to be established by and in sense-experience. On this interpretation, Moore’s procedure is perfectly in order. Nevertheless Carnap would agree with Stroud that Moore’s procedure is powerless to answer the *philosophical* question whether there really are physical things, powerless to establish the *philosophical* proposition that there really are such things. For Carnap accepts the point that, as the skeptic understands, or, more precisely, as he claims to understand, the words “There exist physical things,” Moore’s experience, or any experience, could be just the way it is without these words expressing a truth; and hence that no course of experience could establish the proposition these words are taken by the skeptic to express; that it is in principle unverifiable in experience. But the conclusion that Carnap draws is not the skeptical conclusion. The conclusion he draws is that the words, so taken, express no proposition at all; they are deprived of meaning so that the question whether the proposition they express is true or false does not arise. There is no theoretical issue here. There is indeed a practical issue: whether or not to adopt, or persist in, a certain convention, to make, or persist in, the choice of the physical-thing language or framework of concepts for the organization of experience. Given that the choice is made, the convention is adopted, or persisted in, then we have, internally to the adopted framework, a host of empirically verifiable thing-propositions and hence, internally to the framework, the trivial truth that there exist physical things. But the *external*, philosophical question, which the skeptic tries to raise, viz. whether the framework in general *corresponds to reality*, has no verifiable answer and hence makes no sense.

Moore, then, according to Stroud, either misses the point of the skeptical challenge or has recourse to an unacceptable dogmatism, a dogmatic claim to knowledge. Carnap, again according to Stroud, does not altogether miss the point, but seeks to smother or extinguish it by what Stroud finds an equally unacceptable verificationist dogmatism. It is all very well, says Stroud, to declare the philosophical question to be meaningless, but it does *seem* to be meaningful; the skeptical challenge, the

skeptical question, *seem* to be intelligible. We should at least need more argument to be convinced that they were not.

Many philosophers would agree with Stroud, as against Carnap, on this point; and would indeed go further and contend both that the skeptical challenge is perfectly intelligible, perfectly meaningful, and that it can be met and answered by rational argument. Descartes was one such; though his appeal to the veracity of God to underwrite, or guarantee the reliability of, our natural inclination to believe in the existence of the physical world no longer seems very convincing; if it ever did. More popular today is the view that the assumption of the existence of a physical world, of physical things having more or less the characteristics and powers which our current physical theory represents them as having, provides a far better *explanation* of the course of our sensory experience than any alternative hypothesis. Such an assumption puts us in the way of a non-arbitrary, full, detailed, coherent causal account of that experience to an extent which no alternative story comes anywhere near rivalling. It can therefore be judged rational to accept it by the same criteria of rationality as govern our assessment of explanatory theories framed in natural scientific inquiry or empirical inquiries generally. I shall return to this answer later.

Stroud does not discuss this approach in quite the form I have given it; but he does discuss a near relation of it, viz. Quine’s suggestion of what he calls a “naturalized epistemology,” which would address itself to the empirical question of how, from the meager data available to us in experience, we come to form the elaborate structure of our ordinary and scientific beliefs about the world.⁵ Stroud acknowledges that such an enquiry is perfectly legitimate in itself; but, he contends, it leaves the skeptical challenge completely untouched. If it were seen as an attempted *answer* to the philosophical question, it would be, he maintains, in no better position than Moore’s commonsense assertion; merely a “scientific” version or analogue of the latter. We may in the end be convinced that Quine’s legitimate naturalistic question is the only substantial one that confronts us; but if we are to be satisfied that this is so, it must first be shown that there is something radically faulty, radically misconceived, about the skeptical challenge, about regarding what Carnap called the external question as raising a genuine issue. But this, says Stroud, has not so far been

shown, either by Carnap, though he asserted it, or anyone else.

It is at this point that Stroud acknowledges the appeal of a kind of argument which he calls "transcendental." Such arguments typically take one of two forms. A philosopher who advances such an argument may begin with a premise which the skeptic does not challenge, viz. the occurrence of self-conscious thought and experience; and then proceed to argue that a necessary condition of the possibility of such experience is, say, knowledge of the existence of external objects or of states of mind of other beings. Or he may argue that the skeptic could not even raise his doubt unless he knew it to be unfounded; i.e. he could have no use for the concepts in terms of which he expresses his doubt unless he were able to know to be true at least some of the propositions belonging to the class all members of which fall within the scope of the skeptical doubt. Stroud remains dubious of the success of such arguments; presumably for the same reasons as he expounded in an earlier article entitled "Transcendental Arguments."⁶ There he confronts the propounder of such arguments with a dilemma. *Either* these arguments, in their second form, are little more than an elaborate and superfluous screen behind which we can discern a simple reliance on a simple form of verification principle *or* the most that such arguments can establish is that in order for the intelligible formulation of skeptical doubts to be possible or, generally, in order for self-conscious thought and experience to be possible, we must take it, *or believe*, that we have knowledge of, say, external physical objects or other minds; but to establish this falls short of establishing that these beliefs are, or must be, true.

The second horn of the dilemma is perhaps the more attractive in that it at least allows that transcendental argument may demonstrate something about the use and interconnection of our concepts. But if the dilemma is sound, the skeptic's withers are unwrung in any case. (Stroud seems to assume without question that the point of transcendental argument in general is an anti-skeptical point; but the assumption may be questioned, as I shall later suggest.) In either case, according to Stroud, the skeptic is unshaken because he does not deny that we do, and need not deny that we must, employ and apply the concepts in question in experiential conditions which we take to warrant or justify their application. His point is, and remains, that the fulfillment of those conditions is consistent with

the falsity of all the propositions we then affirm; and hence that – failing further argument to the contrary – we cannot be said really to *know* that any such propositions are true.

3 Hume: Reason and Nature

Is there any other way with skepticism which is not a variant on those I have referred to, i.e. is neither an attempt directly to refute it by rational argument drawing on commonsense or theological or quasi-scientific considerations nor an attempt indirectly to refute it by showing that it is in some way unintelligible or self-defeating? I think there is another way. There is nothing new about it, since it is at least as old as Hume; and the most powerful latter-day exponent of a closely related position is Wittgenstein. I shall call it the way of Naturalism; though this name is not to be understood in the sense of Quine's "naturalized epistemology."

In a famous sentence in Book II of the *Treatise* Hume limits the pretensions of reason to determine the ends of action.⁷ In a similar spirit, towards the end of Book I, he limits the pretensions of reason to determine the formation of beliefs concerning matters of fact and existence. He points out that all arguments in *support* of the skeptical position are totally inefficacious; and, by the same token, all arguments *against* it are totally idle. His point is really the very simple one that, whatever arguments may be produced on one side or the other of the question, we simply *cannot help* believing in the existence of body, and *cannot help* forming beliefs and expectations in general accordance with the basic canons of induction. He might have added, though he did not discuss this question, that the belief in the existence of other people (hence other minds) is equally inescapable. Hume regularly expresses his point by reference to Nature, which leaves us no option in these matters but "by absolute and uncontrollable necessity" determines us "to judge as well as to breathe and feel." Speaking of that total skepticism which, arguing from the fallibility of human judgment, would tend to undermine all belief and opinion, he says: "Whoever has taken the pains to refute the cavils of this total scepticism has really disputed without an antagonist and endeavoured by arguments to establish a faculty which Nature has antecedently implanted in the mind and rendered unavoidable."⁸ He goes on to point out that what holds

for total skepticism holds also for skepticism about the existence of body. Even the professed skeptic “*must assent* to the principle concerning the existence of body, though he cannot pretend by any arguments of philosophy to maintain its veracity”; for “nature has not left this to his choice, and has doubtless esteemed it an affair of too great importance to be entrusted to our uncertain reasonings and speculations.” Hence “’tis vain to ask Whether there be body or not? That is a point which we must take for granted in all our reasonings.”⁹

Here I interpolate some remarks which are not strictly to the present purpose but which are very much to the purpose if one is considering the question of Hume himself. Hume contrasts the vain question, *Whether there be body or not?* with a question he says “we may well ask,” viz. *What causes induce us to believe in the existence of body?* – thus seeming to anticipate Quine’s program for a naturalized epistemology. But there follows, in Hume, what seems to be a striking inconsistency between principle and practice. For, having said that the existence of body is a point which we must take for granted in *all* our reasonings, he then conspicuously does *not* take it for granted in the reasonings which he addresses to the causal question. Indeed those reasonings famously point to a skeptical conclusion. So, as he himself is the first to acknowledge,¹⁰ there is an unresolved tension in Hume’s position (a tension which may be found reminiscent in some ways of the tension between Kant’s empirical realism and his transcendental idealism). One might speak of two Humes: Hume the skeptic and Hume the naturalist; where Hume’s naturalism, as illustrated by the passages I quoted, appears as something like a refuge from his skepticism. An exponent of a more thoroughgoing naturalism could accept the question, *What causes induce us to believe in the existence of body?* as one we may well ask, as one that can be referred to empirical psychology, to the study of infantile development; but would do so in the justified expectation that answers to it would in fact take for granted the existence of body.

Hume, then, we may say, is ready to accept and to tolerate a distinction between two levels of thought: the level of philosophically critical thinking which can offer us no assurances against skepticism; and the level of everyday empirical thinking, at which the pretensions of critical thinking are completely overridden and suppressed by Nature, by an inescapable natural commitment to

belief: to belief in the existence of body and in inductively based expectations. (I hinted at a parallel with Kant; and a parallel there is, though it is only a loose one. There is a parallel in that Kant also recognizes two levels of thought: the empirical level at which we justifiably claim knowledge of an external world of causally related objects in space; and the critical level at which we recognize that this world is only appearance, appearance of an ultimate reality of which we can have no positive knowledge at all. The parallel, however, is only a loose one. Where Hume refers to an inescapable *natural disposition* to belief, Kant produces *argument* (transcendental argument) to show that what, at the empirical level, is rightly reckoned as empirical knowledge of an external world of law-governed objects is a necessary condition of self-awareness, of knowledge of our own inner states; and – a yet more striking difference – where, at the critical level, Hume leaves us with unrefuted skepticism, Kant offers us his own brand of idealism.)

Here I end my digression concerning the complex tensions in Hume’s thought and the parallels with Kant; and return to a consideration of Hume as naturalist, leaving on one side Hume the skeptic. According to Hume the naturalist, skeptical doubts are not to be met by argument. They are simply to be neglected (except, perhaps, in so far as they supply a harmless amusement, a mild diversion to the intellect). They are to be neglected because they are *idle*; powerless against the force of nature, of our naturally implanted disposition to belief. This does not mean that Reason has no part to play in relation to our beliefs concerning matters of fact and existence. It has a part to play, though a subordinate one: as Nature’s lieutenant rather than Nature’s commander. (Here we may recall and adapt that famous remark about Reason and the passions.) Our inescapable natural commitment is to a general frame of belief and to a general style (the inductive) of belief-formation. But *within* that frame and style, the requirement of Reason, that our beliefs should form a consistent and coherent system, may be given full play. Thus, for example, though Hume did not think that a rational justification of induction in general was either necessary or possible, he could quite consistently proceed to frame “rules for judging of cause and effect.” Though it is Nature which commits us to inductive belief-formation in general, it is Reason which leads us to refine and elaborate our inductive canons and procedures and, in their light, to

criticize, and sometimes to reject, what in detail we find ourselves naturally inclined to believe.

4 Hume and Wittgenstein

In introducing this way with skepticism, I associated the name of Wittgenstein with that of Hume. I have in mind primarily Wittgenstein's notes *On Certainty*.¹¹ Like Hume, Wittgenstein distinguishes between those matters – those propositions – which are up for question and decision in the light of reason and experience and those which are not, which are, as he puts it, “exempt from doubt.” Of course there are differences between Hume and Wittgenstein. We do not, for example, find in Wittgenstein any explicit repetition of Hume's quite explicit appeal to Nature. But, as we shall see, the resemblances, and even the echoes, are more striking than the differences. Above all, there is, in Wittgenstein's work, as in Hume's, the distinction between “what it is vain” to make a matter of inquiry, what “we must take for granted in all our reasonings,” as Hume puts it, on the one hand, and what is genuinely matter for inquiry on the other.

Wittgenstein has a host of phrases to express this antithesis. Thus he speaks of a kind of conviction or belief as “*beyond being justified or unjustified*; as it were, as something *animal*” (359);¹² and here we may find an echo of Hume's appeal to Nature and, even more, of Hume's remark that “belief is more properly an act of the sensitive than of the cogitative part of our nature.”¹³ Again, Wittgenstein says that “certain propositions seem to *underlie* all questions and all thinking” (415); that “some propositions are *exempt from doubt*” (341); that “certain things are *indeed* [in der Tat, in practice] not doubted” (342); he speaks of “belief that is not founded” (253) but “in the entire system of our language-games *belongs to the foundations*” (411). Again, he speaks of “propositions which have a *peculiar logical role* in the system [of our empirical propositions]” (136); which belong to our “*frame of reference*” (83); which “*stand fast or solid*” (151); which constitute the “world-picture” which is “the *substratum* of all my enquiring and asserting” (162) or “the *scaffolding* of our thoughts” (211) or “the element in which arguments have their life” (105). This world-picture, he says, is not something he has because he has satisfied himself of its correctness. “No: it is the inherited background against which I

distinguish between true and false” (94). He compares the propositions describing this world-picture to the rules of a game which “can be learned purely practically without learning any explicit rules” (95).

Though the general tendency of Wittgenstein's position is clear enough, it is not easy to extract a wholly clear consecutive statement of it from the mass of figures or metaphors which I have illustrated. Evidently his aim, at least in part, is to give a realistic account or description of how it actually is with our human systems or bodies of belief. Evidently, too, he distinguishes, as I have said, between those propositions, or actual or potential elements in our belief-systems, which we treat as subject to empirical confirmation or falsification, which we consciously incorporate in our belief-system (when we do) for this or that *reason* or on the basis of this or that *experience*, or which we actually treat as matter for inquiry or doubt – and, on the other hand, those elements of our belief-system which have a quite different character, alluded to by the figures of scaffolding, framework, background, substratum, etc. (The metaphors include that of foundations; but it is quite clear that Wittgenstein does not regard these propositions, or elements of the belief-system, as foundations in the traditional empiricist sense, i.e. as basic reasons, themselves resting on experience, for the rest of our beliefs. The metaphor of a scaffolding or framework, within which the activity of building or modifying the structure of our beliefs goes on, is a better one.)

Wittgenstein does not represent this distinction between two kinds of element in our belief-systems as sharp, absolute, and unchangeable. On the contrary. And this is just as well in view of some of his examples of propositions of the second class, i.e. of propositions which are “exempt from doubt.” (Writing in 1950–51, he gives as one example the proposition that no one has been very far (e.g. as far as the moon) from the surface of the earth.) It would have been helpful, though probably contrary to his inclinations, if he had drawn distinctions, or indicated a *principle* of distinction, *within* this class. An indication that there are such distinctions to be drawn comes at the end of an extended metaphor (96–9) in which he compares those propositions which are subject to empirical test to the waters moving in a river and those which are not so subject to the bed or banks of the river. The situation is not unchangeable in that there may sometimes be shifts of the bed or

even of the bank. But, he concludes, “The bank of that river consists partly of hard rock, *subject to no alteration or only to an imperceptible one*, partly of sand which now in one place now in another gets washed away or deposited.”

But how close, really, is Wittgenstein to Hume? There are points at which he may seem closer to Carnap. These are the points at which he seems disposed to express his sense of the difference between those propositions which are subject to empirical test and those which form the scaffolding, framework, foundations etc. of our thought (the hard rock of the river bank) by denying to the latter the status of propositions at all – comparing them, as we have seen, to rules “which can be learned purely practically.” Thus he writes at one point: “No such proposition as ‘There are physical objects’ can be formulated” (36); and even that “‘There are physical objects’ is nonsense” (35). But he is not very close to Carnap. Carnap speaks of a practical issue, a choice – a decision to adopt, or to persist in the use of, a certain framework. There is nothing of this in Wittgenstein. “It is not,” he says, “as if we *chose* the game” (317). And elsewhere, though he is dissatisfied with the expression, we find: “I want to say: propositions of the form of empirical propositions, and not only propositions of logic, form the foundation of all operating with thoughts (with language)” (401). (There is here an evident allusion to the *Tractatus*.) Later, straightforwardly enough, we find: “certain propositions seem to underlie all questions and all thinking.” The apparent shilly-shallying over “proposition” is perhaps palliated by the remarks at 319–20, where he speaks of a lack of sharpness in the boundary between rule and empirical proposition and adds that the concept “proposition” is itself not a sharp one.¹⁴

To sum up now the relations between Hume and Wittgenstein. Hume’s position seems much the simpler. All that is explicitly mentioned by him as constituting the framework of all inquiry – what is to be “taken for granted in all our reasoning” – amounts to two things: acceptance of the existence of body and of the general reliability of inductive belief-formation. This is the groundwork; and its source is unambiguously identified. These unavoidable natural convictions, commitments, or prejudices are ineradicably implanted in our minds by Nature. Wittgenstein’s position is, as we have seen, at least superficially more complicated. First, the propositions or crypto-propositions of the framework, though they may be

taken to include the two Humean elements, are presumptively more various. Second, the framework is, up to a point at least, dynamically conceived: what was at one time part of the framework may change its status, may assume the character of a hypothesis to be questioned and perhaps falsified – some of what we would now regard as assumptions about supernatural agents or powers presumably come into this category – whereas other parts of the framework remain fixed and unalterable. Finally, and connectedly, Wittgenstein does not speak, as Hume does, of one exclusive source, viz. Nature, for these *préjugés*. Rather, he speaks of our learning, from childhood up, an activity, a practice, a social practice – of making judgments, of forming beliefs – to which the crypto-propositions have the special relation he seeks to illuminate by the figures of framework, scaffolding, substratum etc.; that is, they are not judgments we actually make or, in general, things we explicitly learn or are taught in the course of that practice, but rather reflect the general character of the practice itself, form a frame within which the judgments we actually make hang together in a more or less coherent way.

In spite of the greater complication of Wittgenstein’s position, we can, I think, at least as far as the general skeptical questions are concerned, discern a profound community between him and Hume. They have in common the view that our “beliefs” in the existence of body and, to speak roughly, in the general reliability of induction are not grounded beliefs and at the same time are not open to serious doubt. They are, one might say, outside our critical and rational competence in the sense that they define, or help to define, the area in which that competence is exercised. To attempt to confront the professional skeptical doubt with arguments in support of these beliefs, with rational justifications, is simply to show a total misunderstanding of the role they actually play in our belief-systems. The correct way with the professional skeptical doubt is not to attempt to rebut it with argument, but to point out that it is idle, unreal, a pretense; and then the rebutting arguments will appear as equally idle; the reasons produced in those arguments to justify induction or belief in the existence of body are not, and do not become, *our* reasons for these beliefs; there is no such thing as *the reasons for which we hold* these beliefs. We simply cannot help accepting them as defining the areas within which the questions come up of what beliefs we should rationally hold on such-and-such

a matter. The point may be underlined by referring again to some attempts to rebut skepticism by argument.

Perhaps the best skepticism-rebutting argument in favor of the existence of body is the quasi-scientific argument I mentioned earlier: i.e., that the existence of a world of physical objects having more or less the properties which current science attributes to them provides the best available explanation of the phenomena of experience, just as accepted theories within physical science supply the best available explanations of the physical phenomena they deal with. But the implicit comparison with scientific theory simply proclaims its own weakness. We accept or believe the scientific theories (when we do) just because we believe they supply the best available explanations of the phenomena they deal with. That is our reason for accepting them. But no one accepts the existence of the physical world *because* it supplies the best available explanation etc. That is no one's reason for accepting it. Anyone who claimed it was his reason would be pretending. It is, as Hume declared, a point we are naturally bound to take for granted in all our reasonings and, in particular, in all those reasonings which underlie our acceptance of particular physical theories.

Similarly, the best argument against other-minds skepticism is, probably, that, given the non-uniqueness of one's physical constitution and the general uniformity of nature in the biological sphere as in others, it is in the highest degree improbable that one is unique among members of one's species in being the enjoyer of subjective states, and of the kind of subjective states one does enjoy in the kind of circumstances in which one enjoys them. But, again, this is no one's reason for believing in the existence of other minds, of other people, subjects of just such a range of sensations, emotions, and thoughts as he is aware of in himself. We simply react to others as to other *people*. They may puzzle us at times; but that is part of so reacting. Here again we have something which we have no option but to take for granted in all our reasoning.

5 "Only Connect": The Role of Transcendental Arguments

Suppose we accept this naturalist rejection both of skepticism and of skepticism-rebutting arguments as equally idle – as both involving a misunder-

standing of the role in our lives, the place in our intellectual economy, of those propositions or crypto-propositions which the skeptic seeks to place in doubt and his opponent in argument seeks to establish. How, in this perspective, should we view arguments of the kind which Stroud calls "transcendental"? Evidently not as supplying the reasoned rebuttal which the skeptic perversely invites. Our naturalism is precisely the rejection of that invitation. So, even if we have a tenderness for transcendental arguments, we shall be happy to accept the criticism of Stroud and others that either such arguments rely on an unacceptably simple verificationism or the most they can establish is a certain sort of interdependence of conceptual capacities and beliefs: e.g., as I put it earlier, that in order for the intelligible formulation of skeptical doubts to be possible or, more generally, in order for self-conscious thought and experience to be possible, we must take it, or *believe*, that we have knowledge of external physical objects or other minds. The fact that such a demonstration of dependence would not refute the skeptic does not worry our naturalist, who repudiates any such aim. But our naturalist might well take satisfaction in the demonstration of these connections – if they can indeed be demonstrated – for their own sake. For repudiation of the project of wholesale validation of types of knowledge-claim does not leave the naturalist without philosophical employment. E. M. Forster's motto – "only connect" – is as valid for the naturalist at the philosophical level as it is for Forster's characters (and us) at the moral and personal level. That is to say, having given up the unreal project of wholesale validation, the naturalist philosopher will embrace the real project of investigating the connections between the major structural elements of our conceptual scheme. If connections as tight as those which transcendental arguments, construed as above, claim to offer are really available, so much the better.

Of course, it is often disputed, both in detail and in general, that arguments of this kind do or can achieve even as much as the most that Stroud allowed them. Typically, a transcendental argument, as now construed, claims that one type of exercise of conceptual capacity is a necessary condition of another (e.g. that taking some experiences to consist in awareness of objects in physical space is a necessary condition of the self-ascription of subjective states as ordered in time or that being equipped to identify some states of mind in others is a necessary condition of being able to ascribe any

states of mind to ourselves). I am not now concerned with the question of the validity of such arguments but with the general character of the criticisms to which they are typically subject. Typically, the criticism is that what is claimed to be a necessary condition has not been shown to be so and could not be shown to be so without eliminating all possible (or candidate) alternatives, a task which is not attempted. The transcendental arguer is always exposed to the charge that even if *he* cannot conceive of alternative ways in which conditions of the possibility of a certain kind of experience or exercise of conceptual capacity might be fulfilled, this inability may simply be due to lack of imagination on his part – a lack which makes him prone to mistake sufficient for necessary conditions.

It is not my present purpose to inquire how successfully arguments of the kind in question (on the present relatively modest construal of their aims) survive these criticisms; to inquire, that is, whether some or any of them are strictly

valid. I am inclined to think that at least some are (e.g. self-ascription implies the capacity for other-ascription), though I must admit that few, if any, have commanded universal assent among the critics. But whether or not they are strictly valid, these arguments, or weakened versions of them, will continue to be of interest to our naturalist philosopher. For even if they do not succeed in establishing such tight or rigid connections as they initially promise, they do at least indicate or bring out conceptual connections, even if only of a looser kind; and, as I have already suggested, to establish the connections between the major structural features or elements of our conceptual scheme – to exhibit it, not as a rigidly deductive system, but as a coherent whole whose parts are mutually supportive and mutually dependent, interlocking in an intelligible way – to do this may well seem to our naturalist the proper, or at least the major, task of analytical philosophy. As indeed it does to me. (Whence the phrase, “descriptive (as opposed to validatory or revisionary) metaphysics.”)

Notes

- 1 In J. H. Muirhead (ed.), *Contemporary British Philosophy*, 2nd series (London: Allen and Unwin, 1925; reprinted in G. E. Moore, *Philosophical Papers* (London: Allen and Unwin, 1959).
- 2 This vol., ch. 2.
- 3 In P. Bieri, R. P. Horstmann, and L. Kruger (eds), *Transcendental Arguments and Science* (Dordrecht: Reidel, 1979).
- 4 Carnap, “Empiricism, Semantics and Ontology,” *Revue Internationale de Philosophie* (1950), vol. 11. Reprinted in L. Linsky (ed.), *Semantics and the Philosophy of Language* (Champaign: University of Illinois Press, 1952).
- 5 N. V. Quine, “Epistemology Naturalized,” this vol., ch. 23; see also *The Roots of Reference* (LaSalle, IL: Open Court, 1973).
- 6 *Journal of Philosophy*, 1968; reprinted in T. Penelham and J. J. MacIntosh (eds), *The First Critique* (Belmont: Wadsworth, 1969) and in Walker (ed.), *Kant on Pure Reason* (Oxford: Oxford University Press, 1982).
- 7 “Reason is and ought only to be the slave of the passions and can never pretend to any other office than to serve and obey them.” *Treatise of Human Nature*, ed. Selby-Bigge, bk. 2, sec. 3, p. 415.
- 8 *Ibid.*, p. 183.
- 9 *Ibid.*, p. 187.
- 10 *Ibid.*, bk. 1, pt. 4, sec. 7, *passim*.
- 11 Wittgenstein, *On Certainty* (Oxford: Basil Blackwell, 1969).
- 12 Each quoted phrase is followed by its paragraph number in the text of *On Certainty*. Italics are generally mine.
- 13 *Treatise*, bk. 1, pt. 4, sec. 1, p. 183. Another Humean echo is found at para. 135: “But we do not simply follow the principle that what has always happened will happen again (or something like it)? What does it mean to follow this principle? Do we really *introduce* it into our reasoning? Or is it merely the *natural law* which our inferring apparently follows? This latter it may be. It is not an item in our considerations.”
- 14 The restrictions which Wittgenstein is conspicuously inclined to place on the concept of knowledge, on the use of the verb “to know,” reflect, and more emphatically, the inclination to restrict the application of the concept of a proposition. Only what are clearly propositions subject to empirical testing are, he consistently implies, proper objects of the verb “to know”; just as only they can be genuinely objects of doubt.

An Argument for Skepticism

Peter Unger

I mean to offer a positive argument for skepticism about knowledge; I do not mean just to raise some doubts, however general, about statements to the effect that people know. The argument to be offered has as its conclusion the universal form of the skeptical thesis, that is, the proposition that nobody ever knows *anything* to be so. If this argument is sound, as I am inclined to think, then it will follow in particular that nobody ever knows anything about the past or future or even the present, about others or even about himself, about external objects or even about his own experiences, about complicated contingencies or even the simplest mathematical necessities. This, then, is an argument for an extremely strong and sweeping conclusion indeed.

The opposite of skepticism is often called dogmatism. In these terms, dogmatism is the view that certain things are known to be so. The stronger the form of dogmatism, the more sorts of things would be claimed to be known and, so, the weaker the form of skepticism which might still be allowed to hold. Thus, one might be a dogmatist about the past but a skeptic about the future in the sense that one might hold that we know a fair amount about the past but know nothing of the future. But typical arguments to the effect that we know things about the past do not *look* dogmatic in any usual sense. And, arguments to the effect that we know nothing of the future do not in any standard sense *look* particularly undogmatic; they do not *look* particularly indicative of an open-minded approach to things. Going by the typical argu-

ments, then, the label "dogmatist" is unfairly prejudicial and there is no force in the claim that skepticism is to be preferred because the alternative is dogmatism. Unlike such typical arguments, the argument I mean to offer gives substance to the claim that the alternative to skepticism is a view which sanctions a dogmatic attitude. In that one may well not appreciate that this is indeed skepticism's only alternative, one might, perhaps, innocently believe that one knows things without being dogmatic in the process. But once the implications of that belief are brought out, as my argument means to do, the persistence in such a belief may itself be considered dogmatic. Of course, I do not want to be dogmatic in asserting any of this and, indeed, confess to only a moderate amount of confidence in what I have to offer. But as I am inclined to think it true, I offer it in a spirit which I hope may be taken as quite undogmatic and open-minded.

1 A Preliminary Statement of the Argument

I begin by giving a statement of the argument which, while correct in all essentials, does not account for certain complications. On this statement, the argument is exceedingly simple and straightforward. It has but two premises and each of them makes no exceptions whatsoever. The first of these is the proposition:

- (1) If someone *knows* something to be so, then it is all right for the person to be absolutely *certain* that it is so.

For example, if it is true that Knute *knows* that there was a general called “Napoleon,” then it is (perfectly) all right for him to be absolutely *certain* that there was. And, if Rene really *knows* that he exists, then it is (perfectly) all right for Rene to be absolutely certain that he does.

Our second and final premise, then, is this categorical proposition:

- (2) It is never all right for anyone to be absolutely *certain* that anything is so.

According to this premise, it is not all right for Knute to be absolutely certain that there was a general called “Napoleon,” nor is it even all right for Rene to be absolutely certain that he exists. No matter what their situations, these people should not have this “attitude of absolute certainty.” When one understands what is involved in having this attitude, or in being absolutely certain of something, one will presumably understand why it is never all right to be absolutely certain.

These two premises together entail our conclusion of universal skepticism:

- (3) Nobody ever *knows* that anything is so.

In particular, Knute does not really *know* that there was a general called “Napoleon,” nor does Rene really *know* that he exists.

2 The First Premise: The Idea that if One Knows it is All Right for One to be Certain

We often have the idea that someone is certain of something but he shouldn't be. Perhaps from his expressive behavior, perhaps from something else, we *take* it that he is certain of something – whether or not he really is certain of it. We ask him, if we are so inclined, “How can you be *certain* of that?” In asking this question, we manage to imply that it might not be all right for him to be certain and imply, further, that this is because he might not really *know* the thing. If the man could show us that he does know, then we should withdraw the question and, perhaps, even apologize for implying what we did by raising it. But, then, how do we manage to imply so much just by asking this question in the first place? Neither “know” nor any cognate expression ever crosses our lips in the asking. We are able to imply so much, I suggest, because we all accept the idea that, at least generally, if one does

know something then it is all right for one to be certain of it – but if one doesn't then it isn't. This suggests that there is some analytic connection between knowing, on the one hand, and on the other, its being all right to be certain.

The very particular idea that knowing *entails* its being all right to be certain is suggested, further, by the fact that knowing entails, at least, that one *is* certain. That this is a fact is made quite plain by the inconsistency expressed by sentences like “He really *knew* that it was raining, but he *wasn't* absolutely *certain* that it was.” Such a sentence can express no truth: if he wasn't certain, then he didn't know. We get further confirmation here from considering transitivity. The sentences “He was *sad* that it was raining, but he *didn't know* it was” and “He was really *sad* that it was raining, but he *wasn't* absolutely *certain* it was” are likewise inconsistent. Their inconsistency means an entailment from being sad that to knowing, in the first case, and to being certain in the second. This can be best explained, it would seem, by an entailment from knowing to being certain. The entailment from knowing to being certain is convincingly clinched, I think, by appreciating the equivalence between someone's knowing something and his knowing it for certain, or with absolute certainty. To be sure, we may describe cases which we would more naturally react to with the words “He knew it” than “He knew it for certain”: Consider a man who, looking for his cuff links, unerringly went to the very spot they were while doubts went through his mind. Did he know that they were in that spot? But our readiness to say he knew might only indicate loose usage of those words by us, while we are more strict in our use when the word “certain” enters the picture. That this is much the more plausible hypothesis than thinking there to be an inequivalence here is evidenced by the inconsistency of the relevant sentences: “He *knew* it, but he *didn't* know it for certain,” “He really *knew* it, but he *didn't* know it *with absolute certainty*,” “He knew it was there, but he *didn't really know* it,” and so on. No truth can be found in these words no matter when they might be uttered. Even if they are put forth at the end of stories like that of the cuff-link finder, where we are inclined at first to say he knows, we realize that they must express what is false. Accordingly, we are forced to be unswayed by our tendency to loose usage and to admit the equivalence between knowing with absolute certainty and just plain knowing to be so. Admitting this equivalence, we can be quite

confident that knowing does indeed entail being absolutely certain.

Now, it cannot be too strongly emphasized that everything I said is meant to be compatible with the sense which the ordinary word "know" actually has. Indeed, it fairly relies on this word's having only one ("strong") sense as it occurs in sentences of the forms "S knows that *p*" and "S knows about X." Some philosophers have suggested "weak" senses of "know" in which it does not even have an entailment to absolute certainty.¹ But though there is some reason to suppose that "know" has different meanings in "John knows that Jim is his friend" and "John knows Jim,"² there appears no reason at all to suppose that "knows" may mean different things as it occurs in the former sentence. Indeed, reason seems to favor the opposite view. If a genuine ambiguous sentence has a meaning on which it is inconsistent, there will generally be one also on which it is consistent. Once the latter meaning is pointed out, this difference is appreciated and felt to be quite striking. Thus, the sentence "John really *types* many things, but he produces symbols *only orally*" has an obvious meaning on which it is inconsistent. But, it may be pointed out that "types" has another sense, which it shares (roughly) with "classifies." Once this is pointed out, the consistent meaning is appreciated, and the effect is a striking one. No similar phenomenon is ever found with the sentence "John really *knows* that he types things, but he *isn't* absolutely *certain* that he does." There may be many *ad hoc* explanations of this fact. But the only plausible explanation is, I think, that "know" doesn't have a weak sense with no entailment to absolute certainty.³

To deny our first premise, then, is to do violence to the meaning of "know" and to our concept of knowledge. If our argument is to be stopped, it must be with the consideration of the second premise. In any case, it is with that premise that the *substantive* claim of the argument is made: It is not only with mere questions of logical relations with which we must now contend. Accordingly, we now come to the largest and most important part of our discussion.

3 What Attitude is Involved in One's being Absolutely Certain?

I will now, at last, begin to argue for the idea that to be absolutely certain of something is, owing to a

certain feature of personal certainty, to be *dogmatic* in the matter of whether that thing is so. It is because of this dogmatic feature that there is always *something* wrong with being absolutely certain. In other words, it is because of this feature that our second premise, (2), is correct. My argument for the idea that this feature ensures this dogmatism falls naturally into two parts. The first part, which will occupy us in this present section, is aimed at specifying the feature. Thus, we will argue here that one's being absolutely certain of something involves one in having a certain severely negative *attitude* in the matter of whether that thing is so: the attitude that *no* new information, evidence or experience which one might ever have will be seriously considered by one to be *at all* relevant to *any possible* change in one's thinking in the matter. The second part is aimed at showing this attitude to be wrongly dogmatic even in matters which may appear to be quite simple and certain. That more normative segment will be reserved for the section immediately to follow.

That such an absolutely severe attitude should be essential to one's knowing is hardly novel with me. Indeed, philosophers who are quite plainly anti-skeptical proclaim just this attitude as essential to one's knowing. Thus Norman Malcolm thinks himself to know that there is an ink-bottle before him, and describes what he takes to be implicit in this knowledge of his:

Not only do I *not have* to admit that (those) extraordinary occurrences would be evidence that there is no ink-bottle here; the fact is that I *do not* admit it. There is nothing whatever that could happen in the next moment that would by me be called *evidence* that there is not an ink-bottle here now. No future experience or investigation could prove to me that I am mistaken. . . .

It will appear to some that I have adopted an *unreasonable* attitude towards that statement. There is, however, nothing unreasonable about it.

In saying that I should regard nothing as evidence that there is no ink-bottle here now, I am not *predicting* what I should do if various astonishing things happened. . . .

That assertion describes my *present* attitude towards the statement that here is an ink-bottle.⁴

Now, Malcolm, it is true, aligns himself with the idea that there are two (or more) senses of "know"

to be found in sentences like “John *knows* that there is an ink-bottle before him.” This idea is neither correct nor essential to his position in those passages. We have already argued, in section 2, that this idea is not correct. That this incorrect idea is not essential to the main thrust of his quoted remarks is, I think, equally clear. For he allows that there is at least *a* sense of “know” where knowing entails one’s having the extreme attitude they characterize. Presumably, that sense, at least, is just the sense where knowing entails being absolutely certain, and the extreme attitude is just the one which is necessarily involved in absolute certainty. In that such philosophers think that when one knows the attitude of certainty is not only present but quite all right, their thinking that the attitude is to be characterized in such severe negative terms is some indirect evidence for thinking so. An attitude which is so *severely* negative as this might well *not* be one which is very often justified. However, even if one wants to avoid skepticism, a concern for the truth about this attitude makes a severe characterization of it quite unavoidable.

The attitude of certainty concerns *any* sequence of experience or events which could consistently be presented to a sentient subject, without its description prejudging the issue on which it might supposedly bear. Thus, one is certain that there is an ink-bottle before one only if one’s attitude is this: Insofar as I care about being right about whether an ink-bottle is or was before me, no matter how things may seem to appear, *I will not count* as contrary evidence even such extraordinary sequences as these:

... when I next reach for this ink-bottle my hand should seem to pass *through* it and I should not feel the contact of any object ... in the next moment the ink-bottle will suddenly vanish from sight ... I should find myself under a tree in the garden with no ink-bottle about ... one or more persons should enter this room and declare that they see no ink-bottle on this desk ... a photograph taken now of the top of the desk should clearly show all of the objects on it except the ink-bottle.⁵

Now, however (nearly) certain one may be that some or all of these sequences will not occur, that is of course not the same thing as being (at all) certain that there is an ink-bottle before one. But, though there are many differences between the

two, perhaps the one which should most clearly be focused on is this: If one is really certain of the ink-bottle, and not just of other things however related, then one’s attitude is that *even if one should* seem to find oneself in *a* contrary garden, one *would disregard* this experience as irrelevant to the question of whether, at the time in question, there is or was an ink-bottle before one. One might resist this characterization, but then, I think, one would lose one’s proper focus on what it is of which one is certain.

Here is a line of resistance to our characterization of being certain. Suppose, in contrast, one’s attitudes were these: *If* strange things seemed to happen, then perhaps I would change my mind, I just might. But, I am absolutely certain that no strange things will ever happen to speak against there being an ink-bottle. Might not these attitudes be those of a man who was *absolutely* certain that *there is an ink-bottle before him*? Might not he be certain of the ink-bottle, not in or by having a completely exclusionary attitude on that matter itself but, rather, indirectly, so to speak, in or by having just such an attitude toward the possibility of apparently contrary appearances?⁶

This suggestion, this line of resistance, is an interesting one, but it is neither correct nor of any use even if it were correct. First, let us notice that *at least almost* invariably when one is even very close to being absolutely certain of something, one is not nearly so certain that no contrary appearances will turn up. For example, you may be quite sure that I am married. But, you will not be quite so sure that no appearances to the contrary might show up: I may be married but say to you “No, I’m not really married. Mary and I don’t believe in such institutions. We only sent out announcements to see the effect – and it’s easier to have most people believe that we are.” I might, at a certain point, say these things to you and get a few other people to say apparently confirmatory things. All of this, and some more if need be, should and would, I think incline you to be at least a bit less certain that I am married. Thus, at least with things where one is *quite* certain, the matter seems to be quite the *opposite* of what was suggested: One will not be so certain that nothing strangely contradictory will turn up – but one will be inclined to reject any such thing even if it does turn up. We may plausibly project that things work quite the same in situations where someone is absolutely certain (if there really are any such).⁷

Let us now take something of which you are as certain as anything, say, that one and one are two. Suppose that you are very sure that your favorite mathematician will never say something false to you about any simple sum. Imagine that he, or God, tells you and insists that one and one are three, and not two. If your attitude is that he is still to be trusted or, at least, that you would no longer be quite so sure of the sum, then you are not absolutely certain that one and one are two. If you think you *are absolutely certain* of this sum, then, I submit, you should think also that your attitude will be to reject entirely the message from the mathematician or God. In this simple arithmetical matter, you are to give it, perhaps unlike other messages from the same source, no weight at all in your thinking. It seems, then, that this line of resistance is not faithful to the idea of being certain of a particular thing. But would it be of any use in countering skepticism, or the skeptic's charge of dogmatism, even if it were right?

It seems to me that it is at least as dogmatic to have the position that it is absolutely certain that nothing will ever even appear to speak against one's position as to have the attitude that any such appearances which might show up should be entirely rejected. What about appearances to the effect that some contrary appearances, their precise nature left open as yet, are likely to show up in the future? If one is absolutely certain that the latter sort of appearances won't ever show up, one would, presumably, have the attitude of rejecting entirely the indication of the former appearances. One's attitude of rejection gets pushed farther back from the matter itself. Perhaps, on our line of resistance, this may go on indefinitely. But each retreat, and the consequent new place for rejection, only makes a man look more and more obvious in his dogmatism and unreasonableness about the whole affair. Even going back no farther than the second level, so to speak, only a quite foolhardy man would, it seems to me, reject out of hand any suggestion that some things might be brought forth to speak against his position. If anything, it is better for him to allow that they may and to be ready to reject them. So, even if our line of resistance had presented us with a case of being certain, the "indirect" way of being certain would hardly help us to avoid the skeptical charge. That is quite surely no way for being perfectly certain to be perfectly all right.

It is important to stress very hard that a clause like "I should regard nothing as evidence that

there is no ink-bottle now" must be regarded as the expression of a man's *current attitude*, and not as any prediction of what he will do under certain future circumstances.⁸ Thus, one may allow that a sentence like the following is indeed consistent: "He is absolutely *certain* that there are automobiles, but he *may* change his mind should certain evidence come up." That is because even if his present attitude is that he will not, things may not happen in accordance with his attitude. For example, things might happen to him which *cause* him to become uncertain. Or, his attitude might just evaporate, so to speak, the new evidence then affecting him in the unwanted way; and so on. Such conditions as these give us a consistent interpretation for the foregoing sentence, even if not a very ordinary one. A sentence which will always express an inconsistency, on the other hand, is obtained once we make sure that our severely negative clause is embedded so that it is clear that the man's *current attitude* is the point. Thus, in contrast with the foregoing, it is always inconsistent to say "He is absolutely *certain* that there are automobiles, but *his attitude* is that he *may* change his mind should certain evidence come up." A proper assessment of the direct linguistic evidence supports the idea that the attitude of certainty is thus absolutely severe.

This direct linguistic evidence cannot be enough to satisfy one that being certain, or the attitude in knowing, demands so much as we claim. And, it is not enough to add the indirect evidence from anti-skeptical authors. What we want is to fit a severe characterization of this attitude into some more general account of things. Toward this end, I now recall my account of absolute terms.⁹ On this account, *absolute adjectives* like "flat," "useless" and "certain" purport to denote a limiting state or situation to which things may approximate more or less closely. Thus, in the case of these adjectives, the modifier "absolutely," as well as "completely" and "perfectly," is redundant apart from points of emphasis. Now, various locutions with "certain" may appear to indicate matters of degree. But they will always admit of a paraphrase where this appearance is dispelled in favor of a more explicit reference to an *absolute limit*: "That's pretty certain" goes into "That's pretty *close to being absolutely* certain"; "He is more certain of this than of that" goes into "He is *absolutely* certain of this but not of that or else he is *closer to being absolutely* certain of this than of that," and so on. None of this is peculiar to "cer-

tain"; the same happens with locutions containing other absolute adjectives. Thus, these sentences seem to denote matters of degree, but their paraphrases dispel the illusory appearance: "That's the flattest (most useless) thing I've ever seen" goes into "That's the only *absolutely* flat (useless) thing I've ever seen or else that's *closer to being absolutely* flat (useless) than anything else I've seen." In light of these paraphrases, we may repose some confidence in the following formula as saying what it is for something to be *x* where that is the same as being absolutely *x*: Something or someone is *x* (flat, useless, certain, etc.) just in case nothing *could possibly* ever be more *x*, or *x*-er, than that thing or person is right now. It is in this strict sense, then, that being certain, and *a fortiori* being absolutely certain, is being at an absolute limit. Now, absolute adjectives typically have contrasting terms which are *relative adjectives*: "certain" has "confident" and "doubtful," "flat" has "bumpy" and "curved," "useless" has "useful" and "serviceable," and so on. Because matters of degree *are* concerned, there is nothing which is deceptive about the locutions with *these* terms: The sentence "He is pretty confident" does not go into the apparently senseless "He is pretty close to being absolutely confident"; nor does "That is very useful" go into "That is very close to being absolutely useful." These relative terms really do denote matters of degree and not any state or situation which is an absolute limit. If something is bumpy, it is *not* true that nothing could possibly be more bumpy or bumpier. And if someone is confident of something, it does not follow that no one could ever be more confident. Now, a necessary condition for the correct application of an absolute adjective is, at least generally, that certain things denoted by relative adjectives be entirely absent. Thus, it is a necessary condition of something's being flat that it be *not at all* bumpy, that is, that bumpiness not be present even in the least degree. Also, it is a necessary condition of being flat that the thing be *not at all* curved, or that curvature or curvedness not be present at all. We might expect the same sort of thing to hold in the case of someone's being certain of something, and indeed it does: If someone is certain of something, then that thing is *not at all* doubtful so far as he is concerned, that is, doubt or doubtfulness is not present at all in that man with respect to that thing. I have already argued this before, but there are other things which must also be entirely absent if a man is to be certain, though their absence

may be included, I suggest, in the absence of all doubt.

One thing which must be entirely absent, and which is, I think, implicit in the absence of all doubt, is this: any *openness* on the part of the man to consider new experience or information as seriously relevant to the truth or falsity of the thing. In other words, if *S* is certain that *p*, then it follows that *S* is *not at all* open to consider any new experience or information as relevant to his thinking in the matter of whether *p*.

4 Why is there Always Something Wrong with having this Absolute Attitude?

At the beginning of his brilliant paper, "Certainty," G. E. Moore, perhaps the most influential opponent of skepticism in this century, makes some assertions and, as he points out, does so in a very positive and definite way. In just this way, he says, for example, that he had clothes on and was not absolutely naked. Moore goes on to note that although he did not expressly *say* of the things which he asserted that he *knew* them to be true, he implied as much by asserting them in the way he did. His words are these:

I *implied* . . . that I myself knew for certain, in each case, that what I asserted to be the case was, at the time I asserted it, in fact the case. And I do not think that I can be justly accused of dogmatism or over-confidence for having asserted these things positively in the way that I did. In the case of some kinds of assertions, and under some circumstances, a man can be justly accused of dogmatism for asserting something positively. But in the case of assertions such as I made, made under the circumstances under which I made them, the charge would be absurd.¹⁰

I think that we may take it that, according to Moore, the reason he could not so be accused is that he was *not* dogmatic here. And the reason for that is that he *knew* these things, e.g., that he was not naked, so that he was *justified* in being absolutely *certain* of them. And, so, in those innocuous circumstances of speech, he was justified in acting out of, or in accord with, his position or attitude of personal certainty. Moore was saying, in effect,

that one could have this by now familiarly characterized attitude without any pain of being at all dogmatic in the matter: That no new experience or information will have any effect at all on one's thinking in the matter at hand, in this case, in the matter of whether at the then present time one is absolutely naked or not. Moore's position here is, then, quite of a piece with Malcolm's thought that it is *not at all unreasonable* of him to allow nothing to count as contrary evidence in the matter of whether an ink-bottle is before him. But Moore's point is more particular than Malcolm's, for he notes the *particular way* in which one who is certain might be thought to be unreasonable, or not justified, in his attitude: He might be thought to be such *in that* he is *dogmatic* in the matter. Moore similarly foreshadows, while focusing more clearly on the form of the opposite view, Hintikka's implication that in many matters one is *justified* in disregarding any further information: In situations where one knows, Moore says or implies, one is not at all dogmatic in having just such an absolutely negative position or attitude. It seems, then, that Moore was more sensitive than these other authors to the possibility that *dogmatism* might (almost) always be charged of one who was absolutely certain, even when he might rather plausibly claim to know. Now, it strikes me as oddly unfortunate, in a way, that others who actually spelled out what was involved in being certain, were not so sensitive to this particular charge. For it is, I think, precisely the feature they spell out which makes the charge of dogmatism live and convincing. By the same token however, it is to Moore's credit that, without articulating the key idea, he was able to sense the charge of *dogmatism* as a particular threat to his position, perhaps as the key one. Indeed, in the three full sentences I quoted, he refers to this charge as many times. We may put the substantial question, then, in these words: Was Moore referring to a charge of some real substance, or was he right in contending that (because he knew) there was really nothing to be feared?

We may now, I think, more fairly assess the question of whether in cases where one is absolutely certain, supposing there are any such, one's attitude is dogmatic at least in some degree. In such a case, there may be no relevant inconsistency, there being no disparity between one's tenacity and willingness to risk and infer. And, it may well be that no one will ever disagree with one, or even be much less certain of the thing. For, when one is absolutely certain, as we are supposing, the

matter is likely to be clear-cut. But, even if nothing rubs the wrong way, from within oneself or without, one's attitude in the matter is this: I will not allow *anything at all* to count as evidence against my present view in the matter. The case being clear-cut, this attitude will cause one no trouble nor bring any challenge. But, what is one to think of it anyway, even if no penalty or embarrassment is liable ever to occur? I think that any reflection at all makes it pretty plain that, no matter how certain things may seem, *this* attitude is always dogmatic and one who has it will always be open to that charge even if circumstances mean that he will never be exposed to it.

Now, in order to see more clearly why, even in the apparently most clear-cut and certain matters, there is something wrong with letting nothing count against one's being right, it will *help* to describe some sequences of experience. I do *not* think that such an appreciation of detail is really necessary to gaining conviction that the attitude of certainty is always dogmatic and, providing there are no other considerations in its favor, to be foregone in favor of a more open-minded position. One must favor such an attitude in any case, no matter how certain something seems and no matter how little one is able to imagine what experiences there might be which, should they ever occur, one had best consider seriously and not just disregard. This is the right view in the matter however poor our own imaginations might be. But, the strength of habits to the contrary being so great, it will be a big help if we can succeed in imagining sequences of experience which seem to cry out for serious thought. Even in the cases of things which at first seem quite certain, then, and beyond any possibility of questioning at all, I will strive to be of service by imagining experiences. These described experiences should help one grasp firmly the idea that the attitude of certainty is always dogmatic.

5 Helpful Experiences for Rejecting the Attitude of Certainty

In quoting Malcolm's meditations on himself and his ink-bottle, we looked at some sequences of experiences which, if they occurred, might rightly be considered to have some weight and, accordingly, result in one's not being quite so certain as before that there is or was an ink-bottle before one. Malcolm says he wouldn't take those experiences

as relevant here, that that is his attitude and that all of that is perfectly all right. I would disagree. But, in any event, it seems that one can easily imagine experiences which are more telling in this regard. And, also, with only more difficulty, one can imagine others which are easily more telling.

In respect of the matter of that ink-bottle, there are, it seems to me, all sorts of possible experiences which might cast some doubt. For example, one may be approached by government officials who seem to demonstrate that the object on one's desk is a container of a material to poison the water supply, which somehow found its way out of government hands and into one's home. It was disguised to look like an ink-bottle, but it is seen to have many small structural features essential to such a container of poison but which no ink-bottles have. One might well think, then, that though this object holds ink it is not an ink-bottle but, rather, is something else. Perhaps, then, there never was *an ink-bottle* before one, but only some such other object. It seems, at any rate, that such an experience as this should not be disregarded out of hand no matter what one eventually should come to think about whether an ink-bottle was before one. An attitude which would thus disregard it seems, then, to be a dogmatic one.

The experience just described is, I suppose, less than completely convincing. And, even if it is admitted that the experience does have some weight, it seems easy enough to retreat to other statements which are not thus susceptible to experimental challenge. For example, one may be, instead, absolutely certain that there is before one something which looks like an ink-bottle, or that there is something with a circular top, or whatever the favored things turn out to be. Though the sort of experience just imagined might go against one's being certain that *an ink-bottle* is before one, such a sequence of experience will not go against one's certainty about many other things: that there are automobiles, that there have been automobiles for quite some time now, and that one is not now absolutely naked. To get a more completely convincing case about one ink-bottle, and to begin to get a convincing case for these less susceptible things, one's imagination must work more radically. Descartes was quite well aware of the problem when he imagined his evil demon. We may do well to follow suit, though in a more modern and scientific vein.¹¹

I begin to imagine a more radical sequence of experience by supposing myself to experience a

voice, coming from no definite location, which tells me this, in no uncertain terms: All the experiences I am having, including that of the voice, are artificially induced. Indeed, this has been going on for all of my conscious life and it will continue to do so. The voice tells me of various experiences I have had, some of which I had myself forgotten almost entirely. It then says that scientists accomplish all of this with me; it seems to tell me what they are like, what I am really like and, in great detail, how they manage to bring about these effects in me. To make its case most convincing, the voice says what experiences I will next have, and next after that and, then, after that. First, I will seem to fly off the face of the earth to a planet where the inhabitants worship me because I have only one mouth. After that, I am to come back to earth and seem to find that I have been elected Secretary-Treasurer of the International Brotherhood of Electricians. Finally, if that is not preposterous enough, I will seem to open up my body and find myself stuffed with fried shrimps, even unto the inner reaches of my thighs. Miraculously enough, I experience just these to happen. The experiences are not as in a dream but indistinguishable from what I call the most ordinary waking experiences – except, of course, for the extraordinary content. Nor does this predicted sequence seem to take place in a flash, or in any very brief interval. To mirror what I take as reality, it seems to take a couple of months. After a convincing talk with the voice at the end of this experiential journey, I am left in a blue homogenous field of visual experience, feeling little but wonder, to think over whether an ink-bottle was ever before me, whether there are now or ever were any automobiles, and so on. Of course, the voice has told me that none of these things ever were, and told me why I thought otherwise. What am I to think now?

My attitude toward these imagined experiences is that if they should occur I would be at least somewhat less certain than I now am about these matters. I would be at least somewhat unconfident, even, that I was not naked at the time in question. This is my present attitude. If things would not develop in accord with it, that would be something I can now only hope will not happen. Moreover, I think it pretty plain that this is the attitude which I ought to have and that anyone who held an opposite one would have a dogmatic attitude in these matters. That is, if one's attitude is that these experiences will not be counted as having any

weight at all, one would be dogmatic in these matters.

Now, some people might have the attitude that if these experiences occurred one should think himself to be quite mad or, at least, to have had his capacity for judgment impaired in some damaging way.¹² My own attitude is more open than this. But it should be pointed out that even this attitude of prospective self-defeat is quite compatible with that of lessening one's confidence. One's total attitude, that is, might be that if the imagined experiences really came to pass one would both be less certain that there ever were automobiles and also be inclined to think that one must have become quite mad. All that I am claiming or need to claim is that one ought to have at least the first part of this total attitude or, more precisely, that one ought not to have the opposite attitude that any such experiences will be completely disregarded.

7 Helpful Experiences for the Hardest Cases

In respect of almost any matter, the possibility of certain imagined sequences of experience makes quite a convincing case that one ought not, on pain of dogmatism, have the attitude of absolute certainty. There are, however, two sorts of matters where something more must be said to explain how such experience might help us to appreciate the wrongness of this severe attitude. I treat them in turn, proceeding from the less to the more difficult.

The first and lesser difficulty concerns certain sorts of matters about the past. The most famous of these, due to Russell,¹³ is the matter of whether the world sprang into existence five minutes ago. But the matter of whether oneself has existed for more than a brief moment will pose the problem more clearly so far as sequences of convincing experiences are concerned. The problem may be put like this: If any sequence of experience is to be convincing, it must itself endure for much more than a brief moment. Even in advance of any experiences which might look to show that one has been in existence only for a brief moment, one can and ought to appreciate this fact about the conditions of convincing. Therefore, it is in any case quite all right to have the attitude that no possible experience will be counted as convincing evidence for the claim that one has existed only for

a brief moment. Rather, one may disregard any new experience which purports to be to this effect.

The difficulty with this reasoning is that it doesn't take into account how new experiences might make us view time differently. If our voice told us new things about time, we might not be able to disregard it without ourselves being dogmatic. Suppose that the voice says that one has been brought into existence only a brief moment ago complete with an accurate understanding of how long temporal intervals are. But one is also provided, the voice says, with an appealing consistent web of ostensible memories: to believe that one has experienced the things it seems to one that one has will be, then, only to believe what is false. Now, the recent experiences one indeed has had are, according to the voice, part of a sequence which has gone on only for a brief moment, a billionth of a second, to be quite precise. And, this includes these very messages that even now are coming to one. Though it seems to one that the experiences have been going on for some months, one has in fact been alive for only a brief moment and, indeed, the world of concrete things, including the source of the voice, has existed for less than a minute. In response to these vocal claims one might put forward some relativistic theory of time on which the claims would make no sense and, at any rate, on which they could not possibly be true. But, that would only be to adduce some theory. And, if there is anything scientific about science it is that one should never be too certain of any theory, no matter how beautiful, comprehensive and powerful it may seem. So it seems that, no matter how one might wish to reply, one would do well to allow some influence for such a sequence of experience as the one just imagined. One should have the attitude, at least, that should it occur one will be not quite *so* certain, as one otherwise might be, that one has been alive for more than a brief moment.

The greatest difficulty in finding possible experience a help in abandoning the attitude of certainty comes, I think, in matters where we think that the only possible error must be a "purely verbal" one. This occurs, I take it, with matters of "immediate experience," e.g., with whether one is now experiencing phenomenal blueness or pain. And, it occurs with the "simplest matters of logical necessity," e.g., with whether two is the sum of one and one. Perhaps the most famous case, due to Descartes, is that concerning one's own present moment thinking and existence,

e.g., whether one now exists. Now, some philosophers have found it quite an article of faith to suppose that there might be anything to answer to the word "I." They would think, I suppose, that what one ought to be sure of is that *something* now exists, leaving it quite open, what that thing might be. Even if it is true that in such matters as these, any error must be purely verbal, why shouldn't the possibility of just such an error make the attitude of absolute certainty dogmatic in these very matters? I have never heard anything to convince me of the opposite. It is said that what one believes or is certain of are propositions or, at least, some things that are too abstract to have uncertainty over words interfere with their status. Let us agree at the outset that we understand such attempts to downgrade the effect that words might have. But, nevertheless, ought not the following story about possible experience cause at least some very small doubts to enter one's mind? Again, we have our voice. After going through the sequence of experiences I described before, the voice tells me that I become easily confused about the meanings of certain terms. It says that on occasions, and now is one of them, I confuse the meaning of "exist," a word which means, roughly, "to continue on in the face of obstacles," with the meaning of "persist," a word which, roughly, shares a meaning with the verb "to be." Consequently, in philosophizing, I often say to myself "I exist" and "It seems certain to me that I exist now." And, I then seem to remember that I have never thought otherwise. But, in fact, of course, I am quite a changeable fellow, and so I rarely if ever *exist*. It is true that I *persist*, as everyone does, and I *should* say *this* when I do that philosophizing. No doubt, I will soon change once again and say and think, rightly, that what I do is *persist*. This will then seem certain to me, which is better than its seeming certain to me that what I do is *exist*, since at least the former is something which is *true*. But, it would be far better still if *neither* ever even *seemed* to be absolutely certain. At the very least, the voice concludes, I ought never to *be* certain of these things, no matter how tempting that might be. This is especially true in my case because I am so changeable and, as a consequence, so often and so easily confused.

I have no doubt that many would want to protest to this voice. Some might say that the matter of whether the words "I exist now" express a truth and that of whether I exist now are two utterly different matters. Now, it is very true that these

matters are very different. But, why should that lead anyone to protest what I am saying? What I am saying is just that under certain conditions of experience one ought to become less certain than before that one indeed *exists*, that one thing one does is *exist*. Indeed, one may be in just such an experiential situation even while being quite confident that the words "I exist now" do indeed express a truth. We may suppose, after all, that the voice tells one that one *does* continue on in the face of obstacles, and so one ought to be confident that one exists, as well as that one persists. Now, it *may* be that there is something deeply wrong with any of these vocal suggestions and, so, that one ought never to allow any to affect one's beliefs or attitudes even in the most minimal way. But I can't see how anyone can be absolutely certain that *this* is so. And, suppose that the *voice itself* went through all those matters with you and told you to rest assured that such verbal confusions can get you, and are now getting you, into error here. In that one might experience even this, so far as I can see, one's attitude in any of these matters ought not to be that of absolute certainty. Thus, one ought not, really, be absolutely certain that one now exists, or that something exists, or that one now feels pain, or whatever. Of course, the source of uncertainty we have just uncovered is present in matters which are not so apparently certain or simple. Thus, we may now appreciate a bit more fully why it is at least a bit dogmatic to be certain that there is an ink-bottle before one, that there ever are any automobiles, or that one has existed for more than a brief moment.

As I said earlier, these imagined sequences of experience are only meant to be a help in coming to the idea that being certain involves being dogmatic. Their role is to exemplify some situations where this feature of dogmatism might be brought out. I hope that the sequences I have described have been thus revealing and, so, convincing. But that they be so is hardly essential to making good our claim. For even if the particular experience one is able to imagine does not seem to jeopardize some statement which seems quite certain, one shouldn't be *sure* that there isn't any such sequence – possibly, even one which a human imagination just can't grasp in advance. And, even if there is no sequence of *experience* which ought to make one less certain, *mightn't* there be some other factor information about which ought to give one pause? Perhaps, there are some currently obscure conceptual truths about the nature of thought and

reason, which show how any thinking at all is parasitic on the possibility of error in the case. No matter how comfortable one feels in his philosophy and his view of the world, I can't see how he might properly be *certain* that there is no way that he could possibly be wrong. He cannot properly be certain that he has given a complete accounting of every sort of experience, evidence and information which might possibly exist. For this reason, if for no other, it will be dogmatic of him ever to have the attitude that he will disregard *any* new experience, evidence and information which run counter to what he holds.

This is our case, then, that being certain involves being dogmatic and, so, that there is always at least *something* wrong with being certain. As we noticed, whatever is wrong with this dogmatism may be overridden by other considerations, considerations which are not properly epistemological ones. But, the fact that there is always some dogmatism, whether overridden or not, means that nobody ever knows anything about anything. In this sense, then, dogmatism is the opposite of skepticism, and the necessary presence of dogmatism means that skepticism is really true.

Notes

- 1 For examples, see Norman Malcolm, "Knowledge and Belief," in *Knowledge and Certainty* (Englewood Cliffs, New Jersey: 1963), pp. 62ff; and Jaakko Hintikka, *Knowledge and Belief* (Ithaca, New York: 1962), pp. 18ff.
- 2 For example, Spanish uses the verb "*saber*" to translate the first of these sentences and "*conocer*" to translate the second, and so for various other languages. This evidence is both indirect and inconclusive, but it is *some* evidence anyway.
- 3 Perhaps philosophers who seem to see more senses than I do are using "sense" in a different sense. Or, perhaps more likely, they are inventing a new sense for "sense," so as to use the word to make important distinctions about the meaning of our expressions. But, without being impertinent, I can only request to see some reason for supposing that, even in such a new sense of "sense," our verb "know" has two senses.
- 4 Malcolm, "Knowledge and Belief," pp. 67–8.
- 5 *Ibid.*, p. 67. The introductory clause "Insofar as I care about being right about..." is left out by Malcolm. I think it may be necessary for ruling out certain counterexamples concerning untoward motivations. As it plays no important part in our argument however, I will leave it out from now on.
- 6 Some such line of resistance was suggested to me by Gilbert Harman and also by Michael Lockwood.
- 7 I owe to Saul Kripke the idea that these observations are important to consider for such matters.
- 8 In a footnote on p. 68 of "Knowledge and Belief," Malcolm says that he doesn't think the word "attitude" is very satisfactory. He would rather put

- things, he says there, in terms of some conditional statements about what he would say or think right now if or when he imagines things now as happening. But, actually, this latter suggestion is much the poorer and, indeed, Malcolm's choice of the word "attitude" is quite apt and satisfactory.
- 9 Peter Unger, "A Defense of Skepticism," *The Philosophical Review*, vol. LXXX, no. 2 (April 1971), sections II–IV. In a later issue of this journal, James Cargite replied to the skeptical suggestions in that paper of mine: "In Reply to a Defense of Skepticism," *The Philosophical Review*, vol. LXXXI, no. 2 (April 1972). Perhaps the present paper may be taken as deepening the debate between myself and this critic in a way that would not be possible in a brief and direct rejoinder on my part.
- 10 Moore, "Certainty," this vol., ch. 4.
- 11 For a rather different but quite congenial description of unsettling experiences, see Edward Erwin's "The Confirmation Machine," in *Boston Studies in the Philosophy of Science*, vol. VIII, ed. Roger C. Buck and Robert S. Cohen.
- 12 Malcolm suggests this sort of view in his lecture "Memory and The Past," *Knowledge and Certainty*, p. 201. He considers it in a somewhat different context, being most concerned there with the proposition that the earth has existed for no more than five minutes. I will treat such propositions as that in the section following this one. My thoughts on this view owe something to conversation with Michael Slote.
- 13 Bertrand Russell, *The Analysis of Mind* (New York: The Macmillan Company, 1921), pp. 159–60.



PART II

Defining Knowledge

Introduction

The papers in this section are concerned with the question of how to analyze knowledge. An analysis of a concept, at least, provides informative, necessary, and sufficient conditions for the concept's application.

Edmund Gettier's landmark paper successfully refuted the traditional analysis of knowledge as justified true belief. Through a series of examples, Gettier showed that one could believe what is true and be justified in so believing and yet fail to know. Justified true belief is not sufficient for knowledge.

What changes must be made to the traditional account, then, to escape the Gettier cases? Peter Klein proposes a constraint on any account of knowledge, which he calls the *felicitous-coincidence principle*:

If S's evidence for p and a description of some of the particular circumstances in which S believed that p are such that it would not be reasonable to expect that p is true (based upon S's evidence), then even if p is true, S does not know that p .

Klein offers the following analysis, which he claims satisfies this constraint:

S knows that p iff

1. p is true at t ;
2. S believes that p at t ;
3. p is evident for S at t ;
4. there is no true proposition such that if it became evident to S at t , p would no longer be evident to S.

Gilbert Harman criticizes this sort of account, relying on an example based originally on one given by Keith Lehrer and Thomas Paxson, Jr. Suppose I see Tom steal a library book and this is the testimony I give before the University Judicial Council. As it happens, later that day, after I have left the hearing room, Tom's mother testifies that Tom is thousands of miles away, but that his identical twin, Buck, who might well do such things, is in town. Suppose, further, that Tom's mother is a pathological liar and that this is clear to all in the courtroom. Myself, I know nothing about Tom's mother, brother, or any further testimony. Do I know that Tom stole the book? If Klein is right, the answer would be no. But Harman claims, intuitively, that it is yes.

Klein replies to this sort of example in a footnote, claiming that those who affirm knowledge in this case are committed to rejecting the felicitous-coincidence principle. For there is a description of the subject's case which is such that it wouldn't be reasonable to expect that the subject's belief is correct. The relevant description is obtained by conjoining a description of the form "Tom's mother said that ..." with the true statement that in situations like this mothers are generally reliable. Given this description of the subject's situation, together with a description of the subject's evidence for thinking Tom stole the book, it is not reasonable to expect that p is true.

Harman himself, believing the example a counterexample, provides the following fourth condition for knowledge that p : "One's conclusion that p is not based solely on reasoning that essentially involves false intermediate conclusions." The problem posed by the example of Tom is then

Introduction

addressed as follows. First, reasoning is construed so as to involve a claim about the evidence one does not possess. Yet the construal is not so strong as to require that there be *no evidence whatever* that if known would destroy one's justification. Rather, the claim must be that there is no *undermining evidence* one does not possess. Although Harman admits he cannot provide criteria for distinguishing true propositions that constitute undermining evidence from those that are such that if known would render the person unjustified, he notes that there is an intuitive difference, and that this difference is at work in our judgments about Tom.

In his selection, Harman also outlines a conception of inductive reasoning, which he believes comports with his fourth condition for knowledge. Roughly put, inductive reasoning is reasoning to the best explanation. More accurately, it is a kind of self-referential reasoning to the best explanation. To paraphrase Harman:

One may infer a conclusion *h* from one's evidence *e* only if *h* is based not only on *e* but also the following self-referential intermediate conclusion: *h* is the best explanation for *e* and there is no undermining evidence I do not possess against *this very conclusion*.

Robert Nozick argues that the key to knowledge is tracking the truth. One knows through perception that there is a bird on the ledge because one wouldn't believe this if it weren't so. Moreover, one must be such that if things were slightly different and there was a bird on the ledge, one would still believe that this were so. Thus, we obtain the preliminary account

S knows that *p* iff

1. *p* is true.
2. S believes that *p*.
3. If *p* were not true, S would not believe *p*.
4. If *p* were true, S would believe that *p*.

Later, the account is modified to account for certain problem cases, e.g., cases in which, although one knows that *p*, had *p* not been the case, one would still have believed that *p*, but for different reasons. Suppose I am at a meeting with my colleague, a meeting that I firmly believe is essential

to the department. My colleague knows of this conviction but doesn't share it. At the meeting, I know through perception that the meeting is taking place today. Yet, it could well be that, had the meeting been called off, I would still have believed it was taking place today. For my colleague would have told me that there was no need for either of us to go. To fix such problems, Nozick invokes the notion of a *method* of belief, a way of coming to believe. Thus, we arrive at the following account:

S knows that *p* iff there is a method *M* such that

1. S knows that *p*, via *M*
2. All other methods *M*₁ via which S believes that *p* but via which S does not know are outweighed by *M*.

The notion of knowing via a method is then explained in terms of tracking:

S knows that *p* via *M* iff

1. *p* is true
2. S believes, via method *M*, that *p*
3. If *p* weren't true and S were to use *M* to arrive at a belief whether *p*, then S wouldn't believe, via *M*, that *p*.
4. If *p* were true and S were to use *M* to arrive at a belief whether *p*, then S would believe, via *M*, that *p*.

One of the great advantages of the account is that, if correct, it would defuse arguments for skepticism. For if Nozick is right about knowledge, it is not closed under known logical implication. There are cases in which one tracks the truth of *p*, tracks the truth of $\langle p \text{ entails } q \rangle$ but fails to track the truth of *q*. The standard skeptical cases are prime examples: I track the truth of both $\langle I \text{ have hands} \rangle$ and $\langle \langle I \text{ have hands} \rangle \text{ entails } \langle I \text{ am not a brain in a vat deceived into thinking I have hands} \rangle \rangle$, but I don't track the truth of $\langle I \text{ am not a brain in a vat deceived into thinking I have hands} \rangle$. For had I been a brain in a vat so deceived, I would still believe I have hands, for my evidence would be just what it is. Thus, I know that I have hands, even though I don't know that I am not a brain in a vat deceived into thinking I have hands.

Further Reading

- Alston, William P., *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- Chisholm, Roderick, *Theory of Knowledge* (Englewood Cliffs: Prentice-Hall, 1966; 2nd edn 1977; 3rd edn 1989).
- Craig, Edward, *Knowledge and the State of Nature* (Oxford: Clarendon Press, 1990).
- Goldman, Alvin, "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1967), pp. 357–72.
- , *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Lehrer, Keith, *Theory of Knowledge* (Boulder, CO: Westview Press, 1990).
- Moser, Paul K., *Knowledge and Evidence* (Cambridge: Cambridge University Press, 1989).
- Plantinga, Alvin, *Warrant and Proper Function* (Oxford: Oxford University Press, 1993).
- Pollock, John, *Contemporary Theories Of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986).
- Roth, M. D. and Galis, L. (eds), *Knowing: Essays in the Analysis of Knowledge* (New York: Random House, 1970).
- Shope, Robert K., *The Analysis of Knowing* (Princeton: Princeton University Press, 1983).
- Sosa, Ernest, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991), Part I.
- Zagzebski, Linda Trinkhaus, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press, 1996).

Is Justified True Belief Knowledge?

Edmund Gettier

Various attempts have been made in recent years to state necessary and sufficient conditions for someone's knowing a given proposition. The attempts have often been such that they can be stated in a form similar to the following.¹

- (a) S knows that P IFF (i) P is true,
 (ii) S believes that P , and
 (iii) S is justified in believing that P

For example, Claisholm has held that the following gives the necessary and sufficient conditions for knowledge.²

- (b) S knows that P IFF (i) S accepts P ,
 (ii) S has adequate evidence for P , and
 (iii) P is true.

Ayer has stated the necessary and sufficient conditions for knowledge as follows.³

- (c) S knows that P IFF (i) P is true,
 (ii) S is sure that P is true, and
 (iii) S has the right to be sure that P is true.

I shall argue that (a) is false in that the conditions stated therein do not constitute a *sufficient* condition for the truth of the proposition that S knows that P . The same argument will show that (b) and

(c) fail if "has adequate evidence for" or "has the right to be sure that" is substituted for "is justified in believing that" throughout.

I shall begin by noting two points. First, in that sense of "justified" in which S's being justified in believing P is a necessary condition of S's knowing that P , it is possible for a person to be justified in believing a proposition which is in fact false. Second, for any proposition P , if S is justified in believing P and P entails Q and S deduces Q from P and accepts Q as a result of this deduction, then S is justified in believing Q . Keeping these two points in mind, I shall now present two cases in which the conditions stated in (a) are true for some proposition, though it is at the same time false that the person in question knows that proposition.

Case I

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

- (d) Jones is the man who will get the job, and Jones has ten coins in his pocket.

Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (d) entails:

- (e) The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e) and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not *know* that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.

Case II

Let us suppose that Smith has strong evidence for the following proposition:

(f) Jones owns a Ford.

Smith's evidence might be that Jones has at all times in the past within Smith's memory owned a car, and always a Ford, and that Jones has just offered Smith a ride while driving a Ford. Let us imagine, now, that Smith has another friend, Brown, of whose whereabouts he is totally ignor-

ant. Smith selects three place names quite at random and constructs the following three propositions:

- (g) Either Jones owns a Ford, or Brown is in Boston.
- (h) Either Jones owns a Ford, or Brown is in Barcelona.
- (i) Either Jones owns a Ford, or Brown is in Brest-Litovsk.

Each of these propositions is entailed by (f). Imagine that Smith realizes the entailment of each of these propositions he has constructed by (f), and proceeds to accept (g), (h), and (i) on the basis of (f). Smith has correctly inferred (g), (h), and (i) from a proposition for which he has strong evidence. Smith is therefore completely justified in believing each of these three propositions. Smith, of course, has no idea where Brown is.

But imagine now that two further conditions hold. First, Jones does *not* own a Ford, but is at present driving a rented car. And second, by the sheerest coincidence, and entirely unknown to Smith, the place mentioned in proposition (h) happens really to be the place where Brown is. If these two conditions hold, then Smith does *not* know that (h) is true, even though (i) (h) is true, (ii) Smith does believe that (h) is true, and (iii) Smith is justified in believing that (h) is true.

These two examples show that definition (a) does not state a *sufficient* condition for someone's knowing a given proposition. The same cases, with appropriate changes, will suffice to show that neither definition (b) nor definition (c) do so either.

Notes

- 1 Plato seems to be considering some such definition at *Theaetetus* 201, and perhaps accepting one at *Meno* 98.
- 2 Roderick M. Chisholm, *Perceiving: A Philosophical*

Study (Ithaca, NY: Cornell University Press, 1957), p. 16.

- 3 A. J. Ayer, *The Problem of Knowledge* (London: Pelican, 1976).

A Proposed Definition of Propositional Knowledge

Peter Klein

The development of a satisfactory definition of propositional knowledge is essential if an adequate theory of knowledge is to become possible. This task becomes all the more urgent because the attempt to develop such a definition of propositional knowledge within the traditional threefold conditions (true, evident or justified, belief) has recently been seriously challenged, and every subsequent attempt to meet this challenge has failed.¹

I wish to put forth a definition of propositional knowledge and defend it. I think this definition does make a satisfactory epistemology possible because it remains neutral in the conflict between various rival epistemological theories.² Neutrality is an essential feature, for if one theory were to claim that a certain type of proposition is not a proper object of knowledge, and another theory were to argue that such a proposition is a proper object of knowledge, both theories must mean the same thing by propositional knowledge if their disagreement is to be genuine.

Although I believe the proposed definition does accomplish what such a definition should, it does not do everything that some may have wished. It cannot encompass all our uses of "S knows that p ," simply because that expression functions in so many various ways.

The relevant use that I am seeking to define occurs in the following paragraph quoted from *Theory of Knowledge* by Chisholm:

In Plato's dialogue, the *Meno*, Socrates remarks: "That there is a difference between right opinion and knowledge is not at all a conjecture with me but something I would particularly assert that I know. There are not many things of which I would say that, but this one, at any rate, I will include among those that I know." [97C] The distinction would seem to be obvious. If one has knowledge, then one also has right or true opinion. But the converse is not true: one may have right or true opinion without having knowledge. Thus, we may guess correctly today, and therefore, have true opinion, but not know until tomorrow. Or we may have true opinion and never know at all. (p. 5)

Ernest Sosa quotes Russell as saying:

It is very easy . . . to give examples of true beliefs that are not knowledge. There is the man who looks at a clock which is not going though he thinks it is and who happens to look at it at the moment when it is right; this man acquires a true belief about the time of the day but cannot be said to have knowledge ("Propositional Knowledge," 33/4)

I cite these examples not only to illustrate the relevant use of "S knows that p ," but also to underscore the point that propositional knowledge must not be equated with accidentally correct belief.

The traditional analysis of this relevant concept of knowledge is:

Originally published in *The Journal of Philosophy* 68, 16 (1971), pp. 471-82.

S knows that p iff

- (1) p is true,
- (2) S believes p ,
- (3) p is evident to S.

There are counterexamples to the traditional definition, but before discussing them it is necessary to make a few comments on the third condition of the traditional definition of knowledge. Chisholm says that p is evident to S if (a) it is more reasonable for S to believe p than to withhold belief in p , and (b) there is no proposition i such that it is more reasonable for S to believe i than it is for him to believe p .³ The second condition is necessary in order to distinguish evident propositions from reasonable ones. It would be a reasonable but not evident belief that the next item I pick from my pocket will be a Lincoln-head penny if I believe that I have only fifty items in my pocket, forty-nine Lincoln-head pennies and one Indian-head penny. Among the evident propositions would be "the probability is .98 that the next coin will be a Lincoln-head penny" and "the next item will be roughly round and copper-colored."⁴

A proposition would be justified or evident for S if the standard warranting criteria held and if S had no reason to believe that the situation was abnormal. For example, S would be justified in believing that it was 5 p.m. if his watch indicated that time, he knew that it had been reliable in the past, and he remembered winding it recently and it appeared to be running. But suppose that it is 5 p.m., but also that the watch had not worked for 24 hours (the second hand moved when S turned his wrist to look at the watch); S would not know that it was 5 p.m. even though that belief was true and evident.

In order to illuminate the nature of this counterexample, consider a case in which S has a true, but not evident, belief. Suppose S believes, correctly, that the card which he has not examined and which he just picked at random from a full deck of cards is the six of spades. He believes that because the last time he selected the six of spades. His true belief that it is the six of spades is not certifiable as knowledge because, given his evidence and the way in which he picked the card and the fact that there are fifty-two cards in a deck, it could just as easily have been any other card. The evidence S had is not sufficient to "grant entitlement" to knowledge, as Sosa puts it.⁵

For the same reason the example just mentioned is a counterexample to the definition of knowledge as true, evident belief. It is merely a coincidence that S's belief is correct that it is 5 p.m. Of course normally, the evidence S has would be sufficient to certify his belief as knowledge, and he would be entitled to claim that he knew it was 5 p.m. I will return later to the distinction between those times when S is entitled to claim that he knows and those times when he does know. What is important here is that in these special circumstances it is merely a felicitous coincidence that his belief is correct. This could be formulated in terms of a general principle which I will call the *felicitous-coincidence principle*: if S's evidence for p and a description of some of the particular circumstances in which S believes that p are such that it would not be reasonable to expect that p is true (based upon S's evidence), even if p is true, S does not know p . Consequently, we might tentatively assert that S's evidence for his belief that p is not sufficiently strong to certify his belief as knowledge if there is some fact which, were S to become aware of it, ought to cause S to retract his knowledge claim.

It may be thought that the felicitous-coincidence principle could be satisfied by restricting the set of propositions that renders p evident. The accidentally correct, evident beliefs would become uncertifiable as knowledge because of some defect in these propositions. For example, it could be stipulated that the set may neither contain a false proposition nor render evident any false proposition.⁶ This surely is an improvement; it would dispose of the counterexample developed above because that set does render evident the false proposition that the watch has been running since it was last wound.

There are, however, additional counterexamples to this strengthened definition which reveal that the restrictions remain too weak. Ernest Sosa ("Two Conceptions of Knowledge") develops one which I will modify. Suppose that S has been working in an office next to Tom Grabit's office for many years and has often spoken informally with Tom, but does not know anything at all about Tom's personal life. One day he sees what he takes to be Tom stealing a library book. S would be justified in believing that Tom did steal the book and he is correct. But, unbeknown to S, Tom has an identical twin who was in town in the library on the day in question. Further, Tom has never stolen a book from the library and Buck, Tom's twin, is a kleptomaniac who steals books

S knows that p iff

- (1) p is true,
- (2) S believes p ,
- (3) p is evident to S.

There are counterexamples to the traditional definition, but before discussing them it is necessary to make a few comments on the third condition of the traditional definition of knowledge. Chisholm says that p is evident to S if (a) it is more reasonable for S to believe p than to withhold belief in p , and (b) there is no proposition i such that it is more reasonable for S to believe i than it is for him to believe p .³ The second condition is necessary in order to distinguish evident propositions from reasonable ones. It would be a reasonable but not evident belief that the next item I pick from my pocket will be a Lincoln-head penny if I believe that I have only fifty items in my pocket, forty-nine Lincoln-head pennies and one Indian-head penny. Among the evident propositions would be "the probability is .98 that the next coin will be a Lincoln-head penny" and "the next item will be roughly round and copper-colored."⁴

A proposition would be justified or evident for S if the standard warranting criteria held and if S had no reason to believe that the situation was abnormal. For example, S would be justified in believing that it was 5 p.m. if his watch indicated that time, he knew that it had been reliable in the past, and he remembered winding it recently and it appeared to be running. But suppose that it is 5 p.m., but also that the watch had not worked for 24 hours (the second hand moved when S turned his wrist to look at the watch); S would not know that it was 5 p.m. even though that belief was true and evident.

In order to illuminate the nature of this counterexample, consider a case in which S has a true, but not evident, belief. Suppose S believes, correctly, that the card which he has not examined and which he just picked at random from a full deck of cards is the six of spades. He believes that because the last time he selected the six of spades. His true belief that it is the six of spades is not certifiable as knowledge because, given his evidence and the way in which he picked the card and the fact that there are fifty-two cards in a deck, it could just as easily have been any other card. The evidence S had is not sufficient to "grant entitlement" to knowledge, as Sosa puts it.⁵

For the same reason the example just mentioned is a counterexample to the definition of knowledge as true, evident belief. It is merely a coincidence that S's belief is correct that it is 5 p.m. Of course normally, the evidence S has would be sufficient to certify his belief as knowledge, and he would be entitled to claim that he knew it was 5 p.m. I will return later to the distinction between those times when S is entitled to claim that he knows and those times when he does know. What is important here is that in these special circumstances it is merely a felicitous coincidence that his belief is correct. This could be formulated in terms of a general principle which I will call the *felicitous-coincidence principle*: if S's evidence for p and a description of some of the particular circumstances in which S believes that p are such that it would not be reasonable to expect that p is true (based upon S's evidence), even if p is true, S does not know p . Consequently, we might tentatively assert that S's evidence for his belief that p is not sufficiently strong to certify his belief as knowledge if there is some fact which, were S to become aware of it, ought to cause S to retract his knowledge claim.

It may be thought that the felicitous-coincidence principle could be satisfied by restricting the set of propositions that renders p evident. The accidentally correct, evident beliefs would become uncertifiable as knowledge because of some defect in these propositions. For example, it could be stipulated that the set may neither contain a false proposition nor render evident any false proposition.⁶ This surely is an improvement; it would dispose of the counterexample developed above because that set does render evident the false proposition that the watch has been running since it was last wound.

There are, however, additional counterexamples to this strengthened definition which reveal that the restrictions remain too weak. Ernest Sosa ("Two Conceptions of Knowledge") develops one which I will modify. Suppose that S has been working in an office next to Tom Grabit's office for many years and has often spoken informally with Tom, but does not know anything at all about Tom's personal life. One day he sees what he takes to be Tom stealing a library book. S would be justified in believing that Tom did steal the book and he is correct. But, unbeknown to S, Tom has an identical twin who was in town in the library on the day in question. Further, Tom has never stolen a book from the library and Buck, Tom's twin, is a kleptomaniac who steals books

quite often. Although S has a true and evident belief that Tom stole the book, it can hardly be certified as knowledge, because, given S's evidence and the particular circumstances, it is simply a lucky coincidence that he is correct. The felicitous-coincidence principle is at work here.

This is a counterexample to the improved traditional definition, because in this case there is no false proposition rendered evident to S.⁷ That is, S has no reason to believe anything at all about Buck. It is not even reasonable for him to believe that Tom does not have a twin brother, although it would be reasonable for him to believe that it is highly probable that Tom does not have a twin brother. But the latter is not false.

The improved definition remains too weak because it is concerned only with the defeat of a knowledge claim by those false propositions rendered evident to S; whereas there are occasions when S's true evident belief fails to be knowledge for reasons that S had no way of anticipating. To return to the felicitous-coincidence principle, if there is *any* circumstance such that, given S's evidence for *p*, it is not reasonable to expect that *p* is true (given S's evidence), even if *p* is true, S does not know that *p*. The circumstances mentioned in the principle need not be circumstances about which S has any evident beliefs. To put the principle in a slightly different manner: If there is any true proposition *d* such that it and S's evidence for *p* would make it unreasonable to expect that *p* is true, S does not know *p*.

Based upon these considerations, the definition I propose is as follows:

S knows that *p* at *t*₁ if and only if

- (i) *p* is true;
- (ii) S believes *p* at *t*₁;
- (iii) *p* is evident to S at *t*₁;
- (iv) there is no true proposition such that if it became evident to S at *t*₁ *p* would no longer be evident to S.

The first three conditions (i)–(iii) have been called the “traditional” conditions, and I will continue so to refer to them. In what follows I will assume that they are necessary conditions of knowledge. For the sake of simplicity I will refer to a true proposition such that if it became evident to S, *p* would no longer be evident to S as a *disqualifying proposition*.

Now several points about the proposed definition become immediately obvious. First, (i) follows

from (iv) and is therefore no longer required as a condition. For if *p* were false, there would be a disqualifying proposition, namely $\sim p$, and hence, if S knows that *p*, *p* must be true. In fact, if one assumes that S believes those things which are evident, (ii) and (iii) are implied by (iv), for (iv) asserts that *p* is already evident to S, and, if it were, S would believe *p*. But for the sake of the argument in this paper, I will continue to use (i), (ii), and (iii) as separate conditions of propositional knowledge because it is condition (iv) that is probably most suspect, and in defending it I cannot assume what follows from it.

A few more comments about the third condition. In spite of its ugliness I will use the expression “the evidency of *p*.” A proposition is evident to S at *t*₁ iff it is more reasonable for S to believe *p* at *t*₁ (given his evidence for it) than to withhold belief in *p* and there is no more reasonable proposition for S to believe at *t*₁. A proposition may be evident yet false. It may be evident to S₁ but fail to be evident to S₂, because S₁ knows something that S₂ does not know, for example. In that sense evidency is person-relative, but it is person-neutral in the sense that, whatever makes *p* evident to S₁, that and that alone would make *p* evident to S₂. Evidency cannot be defined in a more specific manner because the conventions of evidency will vary depending upon the nature of *p*.

But what of counterexamples? Let us look first at those which attack the definition for being too weak. None of the counterexamples discussed so far work against this definition, for in each case there is a disqualifying proposition: “The watch is not working now” and “Tom, who has never before stolen a book, has an identical twin, kleptomaniac brother, Buck, who was in the library on the day in question.” In other words, what shows that S does not have knowledge in each of the previous cases is a true proposition describing the circumstances mentioned in the felicitous-coincidence principle. If the proposed conditions of knowledge are too weak, it must be possible for S to fail to know that *p* even though the four conditions are fulfilled. But if *p* is evident to S and yet S does not know that *p*, there must be some true proposition *d* which shows that in *this case* the evidence S has for *p* is insufficient to warrant certification of S's belief as knowledge. The true proposition *d* would disqualify S's belief as knowledge only if it were such as to make *p* no longer evident (given S's evidence alone), and, because of the person-neutral character of evidency, *d* would

be such that if it became evident to S, p would no longer be evident. Hence, there can be no counterexamples to show that the definition is too weak.

A purported counterexample showing that the conditions are too strong would have to show that S knows that p even though there is a disqualifying proposition. However, as a rough and ready reply it could be pointed out that if there were such a disqualifying proposition, p would no longer be evident to the giver of the counterexample, G, and hence the giver of the counterexample would find himself in the absurd position of claiming that S knows that p but he himself does not know that p .

I said a moment ago that this was a rough and ready reply, but the situation is more complicated than I just implied. The above response is valid only if the disqualifying proposition d is a disqualifying proposition for both S and G. If it were, then, of course, if d were evident to G, p would no longer be evident to him, and hence he could no longer claim that S knows that p , implying that he knows that p and, therefore that p is evident to him. But suppose that d is not a disqualifying proposition for both S and G, but only for S. Ought we still assert that S knows that p ? Consider the following case. Suppose that the gas tank of S's car is one-fourth full and that S sees that his gas gauge reads "1/4," remembers that the gauge has been reliable in the past, and consequently believes that the tank is one-fourth full. The three traditional conditions of knowledge are fulfilled. Now let us suppose that the giver of the counterexample, G, knows that the gauge is not working, but that he has sufficiently strong evidence, which S does not have, so that the claim that the tank is one-fourth full remains evident to him. He may have checked the gas in the tank through some other method. Now, since S does not have this additional evidence, the disqualifying proposition d , "the gas gauge is not working properly" would serve as a disqualifying proposition for S but not for G. Now G must maintain in spite of d that S knows that the tank is empty. That is, the giver of the counterexample must believe that there is a disqualifying proposition d , for S, but not for him, which is such that S still knows that the tank is empty. Is G correct?

In some sense of "know," S does know that the tank is one-fourth full. He "knows," but for the wrong reasons. But in the relevant sense, he does not know because he only happens to have a true belief and it is merely an accident (in this case) that his belief is correct. According to the felicitous-

coincidence principle, S does not know that the tank was one-fourth full. Of course any given disqualifying proposition d need not disqualify all of S's evidence as it does in this case in order to make p no longer evident to S. Then d would not make accidental the connection between *all* of S's evidence for p and p , but, if the evidence that d does disqualify is essential for the evidency of p to S, then the connection between the remaining evidence and the truth of p becomes insufficiently strong to certify his belief as knowledge. He would no doubt retract p if d were to become evident to him. If d disqualified an essential part of the evidence for p that S has, S would not know that p . On the other hand, if d does not disqualify an essential part of the evidence S has for p , then it would not be a disqualifying proposition as defined.

In spite of what I have just claimed, namely, that if the disqualifying proposition were such that it disqualified something essential, S would not know that p , it may be thought that there might still be some counterexamples lurking here. I would like to deal with one possible counterexample, the strongest that I know of, in order to reinforce the above quite general argument to show that the definition is not too strong.

Consider Mr Jones, who goes to the house of an acquaintance M for the first time. He sees some flowers on the mantelpiece, and throughout the evening the various guests and M comment on the flowers. Their comments cohere. Later, Mr Jones discovers that M is a magician and delights in fooling his guests by creating some extremely cleverly devised illusions of flowers on a mantel. Suppose further that on the night in question M was not up to his old tricks. In this example p_1 is "there are flowers on the mantel" and d_1 is "M is a magician."⁸

Now, what about this case? Did Jones know p_1 ? First, let me point out that our intuitions are likely to diverge here, because this is so artificial and unusual a case.

We might believe that Jones did know that p_1 . Our argument would run as follows: proposition d_1 does not disqualify p_1 ; it does so only in conjunction with d_1' , where d_1' equals "M was up to his old tricks last night"; d_1' is made evident by d_1 . But d_1' is false; hence the conjunction $d_1' \cdot d_1$ is false, and there is no proposition that is both true and evident and disqualifies p_1 . Hence the definition is upheld, and this is not a counterexample, because

Jones knew p_1 and there is no disqualifying proposition.

This response is incorrect, however, because d_1 does justify a proposition that is both true and evident and does disqualify p_1 , namely, d_1' : " d_1' is highly probable." If it became evident to Jones that it was highly probable that M was up to his old tricks, then it would no longer be evident to him that there were flowers on the mantel. Hence this is not a counterexample, because the fourth condition is not fulfilled and Jones does not know that p_1 .

But suppose that someone were to claim that Jones did know that p_1 even though d_1'' is a disqualifying proposition for p_1 . Would he be correct? I think not; for what d_1'' asserts is that, in this particular case, even though the standard criteria hold, they are not reliable. Therefore this example is not essentially different from many of the cases presented earlier. In those cases the reason why the subject's belief was not certifiable as knowledge was simply that, although the standard warranting criteria held, in those particular cases the criteria were not reliable. The felicitous-coincidence principle again holds.

If one were to insist that Jones did know p_1 , even though there is a disqualifying proposition, one would not only have to reject the felicitous-coincidence principle; one would also be forced to accept the following rather awkward result: Jones did know p_1 before d_1'' became evident, he no longer knew p_1 after it became evident (because p was no longer evident), and then, finally, he knew p_1 again after it became evident that M was not up to his old tricks. A much more plausible rendering of the situation is this: Jones would have been justified in asserting that he knew that p_1 (although he would have been mistaken) before d_1'' became evident to him; after it became evident he would have been justified in asserting that he did not know p_1 (and he would be correct); and finally, after he learned that M was not up to his old tricks, he would be justified in asserting he knew p_1 (and he would be correct). There are many occasions when we are entitled to claim that we know that p , although we later find out we were mistaken, for it would seem that we are entitled to claim that we know that p whenever we believe p and are justified in believing that the standard warranting criteria for judgments like p are fulfilled and are justified in believing that the situation is not abnormal. But of course we may be entitled to claim that we know that p and be mistaken in the

claim, if either the standard criteria do not hold or the particular situation is not such that the standard criteria are sufficient warranting criteria.⁹

It would seem that those who insist that Jones knew that p_1 before d_1'' became evident fail to distinguish between those occasions when S is entitled to claim that he knows that p and those occasions when such a claim is correct. His claim would be correct only when the four necessary conditions of knowledge were fulfilled.

I said earlier that a good definition of propositional knowledge would have to remain neutral with regard to the disputes among rival epistemological theories; and it may be thought that the fourth condition is so strong that it prejudices the issues in favor of one or another form of skepticism. Some may believe that the definition is so strong that:

- I. If the definition were accepted, S could never know that p , because S could never know that the fourth condition was fulfilled.
- II. If the definition were accepted, S would not be warranted in asserting that he knew that p , because S would never be warranted in asserting that the fourth condition held.
- III. If the definition were accepted, it would never be true that S knows that he knows that p because he could never know that the fourth condition held.

Now if any one of these forms of skepticism were implied by the definition, it would lose its neutrality and, hence, would not be acceptable. The definition must allow for the forms of skepticism involved in I, II, and III, but it ought not to imply them.

In reply to I, let me simply point out that the fourth condition is not "S knows that there is no disqualifying proposition." S can know that p without knowing — or for that matter, without even considering — whether there are any disqualifying propositions. The condition merely asserts that his evidence must be such that there are no disqualifying propositions. A person supporting the skeptical position of I-type must show that for some reason there is always a disqualifying proposition for any particular type of proposition which he held was not a possible object of knowledge.

In reply to II, it must be pointed out that S would be warranted in asserting or believing that he knew that p , if it is evident to him that there is

no disqualifying proposition for p . Although that is never beyond any conceivable doubt, on many occasions it would be beyond any reasonable doubt. Those are the occasions mentioned earlier when S is entitled to claim that he knows that p . A skeptic supporting the position involved in II would have to show that there never are such occasions.

Finally, the statement "S knows that he knows that p " would be true whenever the following are true:

- 1' "S knows that p " is true; (see reply to I)
- 2' S believes that he knows that p ;
- 3' "S knows that p " is evident to S; (see reply to II)
- 4' There is no disqualifying proposition for "S knows that p ."

It seems quite clear that these conditions could be fulfilled; or rather, the definition itself does not rule out the possibility that these conditions are fulfilled.

I said earlier that any good definition of knowledge would have to be acceptable to all the rival epistemological theorists, and I have just shown that mine would be acceptable to three forms of skepticism and their rivals. I cannot show that it will be acceptable to all of the other rivals, but I can at least make this claim more plausible if I can show my definition to be acceptable to Descartes and his critics, since the differences there seem to be about as serious as possible.

The Cartesians believe that it is possible to legitimately question what we ordinarily take to be evident claims, for example, that there is now a piece of paper in front of me. These are dubitable, not because we have some evidence against any particular such claim, but because our methods of determining whether such propositions are true or false are themselves in need of confirmation. The anti-Cartesians, on the other hand, argue that the methods are adequate. Now I do not wish to get involved in the particular disputes between

Descartes and his critics. Nor do I wish to detail the claims of each rival epistemology. But I do want to show that, given this proposed definition of propositional knowledge, the issue can be joined and perhaps settled.

Cartesian doubt, in its strong form, must grant that a certain proposition p is evident, given all the standard tests for p , but yet it must maintain that it remains possible to doubt that we know p . Whereas the anti-Cartesians seem to be maintaining that, if p is evident as a result of all the standard tests being applied to p , then it becomes gratuitous to doubt that we know that p .

The issue, then, can be put as follows: Is there a disqualifying proposition for all those propositions which we ordinarily take to be beyond doubt true? Will, for example, any of the following serve as a disqualifying proposition?

1. Perhaps I am dreaming now.
2. I have been deceived before, so perhaps I am being deceived now.
3. Something other than material objects may be causing our perceptions, for example, an evil genius.
4. Perhaps I am mad.

Now, as I mentioned above, I do not propose to get involved at this time in the dispute over the truth of any of the above propositions. What I wish to point out is that the proposed definition of propositional knowledge does not prejudge the issue at all, in favor of either the Cartesians or the anti-Cartesians. In fact, it clarifies the issues by focusing attention on the considerations surrounding the possibility of the existence of such a disqualifying proposition.¹⁰

The definition of propositional knowledge that I have proposed seems to meet all the necessary conditions of any such attempt and, in addition, seems to provide a procedure for evaluating various epistemological theories.

Notes

1 One series of discussions is: Edmund Gettier, "Is Justified True Belief Knowledge?", this vol., ch. 7; Roderick M. Chisholm, *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), p. 23. fn 22; John Pollock, "Chisholm's Definition of Knowledge," *Philosophical Studies* XIX, 5 (October 1968), pp. 72-6. Another series is Ernest Sosa, "Proposi-

tional Knowledge," *Philosophical Studies* XX, 3 (April 1969), pp. 33-43; Keith Lehrer and Thomas Paxson, Jr., "Knowledge: Undeclared Justified True Belief," *The Journal of Philosophy* LXVI, 8 (24 April 1969), pp. 225-37; Sosa, "Two Conceptions of Knowledge," *The Journal of Philosophy* LXVII, 3 (12 Feb. 1970), pp. 59-68.

- 2 By "an epistemological theory" I mean a set of beliefs concerning the types of propositions that can (or cannot) be known.
- 3 Chisholm, *Theory of Knowledge*, p. 22. Implicit in the concept of *evidency* or *justification* used by Chisholm and the other writers concerned with the adequacy of the traditional definition is the notion of a set of propositions with relative degrees of reasonableness *relevant* to some evidence. This concept of justification, although sufficient for elucidating the justification of empirical propositions and adequate for the purposes of this essay, must be amended if analytic propositions are to be counted as evident. At some time I intend to explore the evidency of analytic propositions, but it is the class of nonanalytic propositions that has proved troublesome to the traditional definition of knowledge and it is those difficulties with which this article is concerned.
- 4 I am indebted to the editors of the *Journal of Philosophy* for pointing out a mistake in an earlier version of this paper; I had failed to note the compatibility of its being more reasonable for S to believe p than to withhold belief and its being more reasonable to believe i than p .
- 5 "Two Conceptions of Knowledge," p. 62.
- 6 Two such attempts to improve the traditional definition were made by Sosa, "Propositional Knowledge," and Lehrer and Paxson, "Knowledge."
- 7 It may be thought that there is one false proposition rendered evident by the propositions which render it evident that Tom Grabit stole the book, i.e., S knows that Tom stole the book. If this were the case, not only would it appear to trivialize the improved traditional definition, but also any other proposed fourth condition would fail to explicate the concept of knowledge for the definition would become merely "S knows that p if it is not false that S does not know that p ." However, whatever renders p evident to S would not, by itself, render it evident that he knows p , because it would not render it evident to S that he believes p . That is, a proposition may be evident for S without S believing that it is. If it is evident to S, this means that it would be more reasonable for him to believe than withhold belief; but, of course, he may not believe that p .
- 8 A counterexample similar to this one was suggested to me both by A. J. Ayer and by Marc Cohen.
- 9 Lehrer and Paxson present and dismiss a definition of knowledge similar to the one developed here. They reject the definition because of a counterexample similar to the Jones case.

Suppose I see a man walk into the library and remove a book from the library by concealing it beneath his coat. Since I am sure the man is Tom Grabit, whom I have often seen before when he attended my classes, I report that I know that Tom Grabit has removed the book. However, suppose further that Mrs Grabit, the mother of Tom, has averred that on the day in question Tom was not in the library, indeed, was thousands of miles away, and that Tom's identical twin brother, Buck Grabit, was in the library. Imagine, moreover, that I am entirely ignorant of the fact that Mrs Grabit has said these things. The statement that she has said these things would defeat any justification I have for believing that Tom Grabit removed the book, according to our present definition of defeasibility. Thus, I could not be said to have nonbasic knowledge that Tom Grabit removed the book ("Knowledge," p. 228).

But the situation is not quite that simple. I grant that it appears that this is a case of S's knowing, but I do not grant that the claim "Tom's mother said..." is, by itself, sufficient to disqualify S's knowledge claim. If we couple what Tom's mother said with the proposition "what mothers say in situations like this is generally reliable," then, if the conjunction of the two propositions is a disqualifying proposition, S would not know that Tom stole the book. That is, if Tom's mother said that Buck stole the book and mothers' statements are generally reliable, it is only a felicitous coincidence that S's belief is correct. For if mothers' statements are generally reliable in these situations, it is highly probable that Buck, and not Tom, stole the book and it is merely a lucky coincidence that S's belief is correct because the propositions that render p evident to S are equally compatible with the highly probable denial of p .
- 10 Let me add that the dispute between the Cartesians and the anti-Cartesians could be viewed as a disagreement over whether the evidency of (iii) to S renders (iv) evident to S.

Selections from *Thought*

Gilbert Harman

Knowledge and Probability

The lottery paradox

Some philosophers argue that we never simply believe anything that we do not take to be certain. Instead we believe it to a greater or lesser degree; we assign it a higher or lower "subjective probability." If knowledge implies belief, on this view we never know anything that isn't absolutely certain. That conflicts with ordinary views about knowledge, since our degree of belief in some things we think we know is greater than our degree of belief in other things we think we know.

We might count as believed anything whose "subjective probability" exceeds .99. But that would also conflict with ordinary views. We do not suppose that a man inconsistently believes of every participant in a fair lottery that the participant will lose, even though we suppose that the man assigns a subjective probability greater than .99 to each person's losing. If ordinary views are to be preserved, belief must be distinguished from high degree of belief.

A rule of inductive inference is sometimes called a "rule of acceptance," since it tells us what we can accept (i.e., believe), given other beliefs, degrees of belief, etc. A purely probabilistic rule of acceptance says that we may accept something if and only if its probability is greater than .99. Kyburg points out that such a rule leads to a "lottery paradox" since it authorizes the acceptance of an inconsistent set of beliefs, each saying of

a particular participant in a lottery that he will lose.¹

It is true that no contradiction arises if conclusions are added to the evidence on whose basis probabilities are calculated. Concluding that a particular person will lose changes the evidential probability that the next person will lose. When there are only 100 people left, we cannot infer the next person will lose, since the evidential probability of this no longer exceeds .99. But this does not eliminate paradox. The paradox is not just that use of a purely probabilistic rule leads to inconsistent beliefs. It is not obviously irrational to have inconsistent beliefs even when we know that they are inconsistent. It has occasionally been suggested² that a rational man believes that he has at least some (other) false beliefs. If so, it follows logically that at least one thing he believes is false (if nothing else, then his belief that he has other false beliefs); a rational man will know that. So a rational man knows that at least one thing he believes is false. Nevertheless it is paradoxical to suppose that we could rationally believe of every participant in a lottery that he will lose; and it is just as paradoxical to suppose that we could rationally believe this of all but 100 participants in a large lottery.

The lottery paradox can be avoided if a purely probabilistic rule of acceptance is taken to be relevant not to the acceptance of various individual hypotheses but rather to the set of what we accept. The idea is that the probability of the whole set must exceed .99. We are free to choose among various hypotheses saying that one or another participant in a lottery loses as long as the probability of the conjunction of all hypotheses accepted

Originally published in G. Harman, *Thought* (Princeton: Princeton University Press, 1973).

remains above .99. (The idea requires a distinction between what is simply accepted and what is accepted as evidence. If we could add new conclusions to the evidence, the lottery paradox would be generated as indicated in the previous paragraph.) However, although this version of a purely probabilistic rule does not yield the lottery paradox, it does not fit in with ordinary views, as I shall now argue.

Gettier examples and probabilistic rules of acceptance

In any Gettier example we are presented with similar cases in which someone infers *h* from things he knows, *h* is true, and he is equally justified in making the inference in either case.³ In the one case he comes to know that *h* and in the other case he does not. I have observed that a natural explanation of many Gettier examples is that the relevant inference involves not only the final conclusion *h* but also at least one intermediate conclusion true in the one case but not in the other. And I have suggested that any account of inductive inference should show why such intermediate conclusions are essentially involved in the relevant inferences. Gettier cases are thus to be explained by appeal to the principle

P Reasoning that essentially involves false conclusions, intermediate or final, cannot give one knowledge.

It is easy to see that purely probabilistic rules of acceptance do not permit an explanation of Gettier examples by means of principle *P*. Reasoning in accordance with a purely probabilistic rule involves essentially only its final conclusion. Since that conclusion is highly probable, it can be inferred without reference to any other conclusions; in particular, there will be no intermediate conclusion essential to the inference that is true in one case and false in the other.

For example, Mary's friend Mr Nogot convinces her that he has a Ford. He tells her that he owns a Ford, he shows her his ownership certificate, and he reminds her that she saw him drive up in a Ford. On the basis of this and similar evidence, Mary concludes that Mr Nogot owns a Ford. From that she infers that one of her friends owns a Ford. In a normal case, Mary might in this way come to know that one of her friends owns a Ford. However, as it turns out in this case, Mary is

wrong about Nogot. His car has just been repossessed and towed away. It is no longer his. On the other hand, Mary's friend Mr Havit does own a Ford, so she is right in thinking that one of her friends owns a Ford. However, she does not realize that Havit owns a Ford. Indeed, she hasn't been given the slightest reason to think that he owns a Ford. It is false that Mr Nogot owns a Ford, but it is true that one of Mary's friends owns a Ford. Mary has a justified true belief that one of her friends owns a Ford but she does not know that one of her friends owns a Ford. She does not know this because principle *P* has been violated. Mary's reasoning essentially involves the false conclusion that Mr. Nogot owns a Ford.⁴

But, if there were probabilistic rules of acceptance, there would be no way to exhibit the relevance of Mary's intermediate conclusion. For Mary could then have inferred her final conclusion (that one of her friends owns a Ford) directly from her original evidence, all of which is true. Mr Nogot *is* her friend, he *did* say he owns a Ford, he *did* show Mary an ownership certificate, she *did* see him drive up in a Ford, etc. If a purely probabilistic rule would permit Mary to infer from that evidence that her friend Nogot owns a Ford, it would also permit her to infer directly that one of her friends owns a Ford, since the latter conclusion is at least as probable on the evidence as the former. Given a purely probabilistic rule of acceptance, Mary need not first infer an intermediate conclusion and then deduce her final conclusion, since by means of such a rule she could directly infer her final conclusion. The intermediate conclusion would not be essential to her inference, and her failure to know that one of her friends owns a Ford could not be explained by appeal to principle *P*.

A defender of purely probabilistic rules might reply that what has gone wrong in this case is not that Mary *must* infer her conclusion from something false but rather that, from the evidence that supports her conclusion, she *could* also infer something false, namely that Mr Nogot owns a Ford. In terms of principle *P*, this would be to count as essential to Mary's inference any conclusion the probabilistic rule would authorize from her starting point. But given any evidence, some false conclusion will be highly probable on that evidence. This follows, e.g., from the existence of lotteries. For example, let *s* be a conclusion saying under what conditions the New Jersey State Lottery was most recently held. Let *q* say what ticket won the

grand prize. Then consider the conclusion, *not both s and q*. Call that conclusion *r*. The conclusion *r* is highly probable, given evidence having nothing to do with the outcome of the recent lottery, but *r* is false. If such highly probable false conclusions were always considered essential to an inference, Mary could never come to know anything.

The problem is that purely probabilistic considerations do not suffice to account for the peculiar relevance of Mary's conclusion about Nogot. Various principles might be suggested; but none of them work. For example, we might suspect that the trouble with *r* is that it has nothing to do with whether any of Mary's friends owns a Ford. Even if Mary were to assume that *r* is false, her original conclusion would continue to be highly probable on her evidence. So we might suggest that an inferable conclusion *t* is essential to an inference only if the assumption that *t* was false would block the inference. That would distinguish Mary's relevant intermediate conclusion, that Nogot owns a Ford, from the irrelevant conclusion *r*, since if Mary assumed that Nogot does not own a Ford she could not conclude that one of her friends owns a Ford.

But again, if there is a purely probabilistic rule of acceptance, there will always be an inferable false *t* such that the assumption that it is false would block even inferences that give us knowledge. For let *h* be the conclusion of any inference not concerned with the New Jersey Lottery and let *r* be as above. Then we can let *t* be the conjunction *h & r*. This *t* is highly probable on the same evidence *e* on which *h* is highly probable; *t* is false; and *h* is not highly probable relative to the evidence *e* & (*not t*). Any inference would be undermined by such a *t*, given a purely probabilistic rule of acceptance along with the suggested criterion of essential conclusions.

The trouble is that purely probabilistic rules are incompatible with the natural account of Gettier examples by means of principle *P*. The solution is not to attempt to modify *P* but rather to modify our account of inference.

Knowledge and Explanation

A causal theory

Goldman suggests that we know only if there is the proper sort of causal connection between our belief

and what we know.⁵ For example, we perceive that there has been an automobile accident only if the accident is relevantly causally responsible, by way of our sense organs, for our belief that there has been an accident. Similarly, we remember doing something only if having done it is relevantly causally responsible for our current memory of having done it. Although in some cases the fact that we know thus simply begins a causal chain that leads to our belief, in other cases the causal connection is more complicated. If Mary learns that Mr Havit owns a Ford, Havit's past ownership is causally responsible for the evidence she has and also responsible (at least in part) for Havit's present ownership. Here the relevant causal connection consists in there being a common cause of the belief and of the state of affairs believed in.

Mary fails to know in the original Nogot-Havit case because the causal connection is lacking. Nogot's past ownership is responsible for her evidence but is not responsible for the fact that one of her friends owns a Ford. Havit's past ownership at least partly accounts for why one of her friends now owns a Ford, but it is not responsible for her evidence. Similarly, the man who is told something true by a speaker who does not believe what he says fails to know because the truth of what is said is not causally responsible for the fact that it is said.

General knowledge does not fit into this simple framework. That all emeralds are green neither causes nor is caused by the existence of the particular green emeralds examined when we come to know that all emeralds are green. Goldman handles such examples by counting logical connections among the causal connections. The belief that all emeralds are green is, in an extended sense, relevantly causally connected to the fact that all emeralds are green, since the evidence causes the belief and is logically entailed by what is believed.

It is obvious that not every causal connection, especially in this extended sense, is relevant to knowledge. Any two states of affairs are logically connected simply because both are entailed by their conjunction. If every such connection were relevant, the analysis Goldman suggests would have us identify knowledge with true belief, since there would always be a relevant "causal connection" between any state of true belief and the state of affairs believed in. Goldman avoids this reduction of his analysis to justified true belief by saying that when knowledge is based on inference

relevant causal connections must be “reconstructed” in the inference. Mary knows that one of her friends owns a Ford only if her inference reconstructs the relevant causal connection between evidence and conclusion.

But what does it mean to say that her inference must “reconstruct” the relevant causal connection? Presumably it means that she must infer or be able to infer something about the causal connection between her conclusion and the evidence for it. And this suggests that Mary must make at least two inferences. First she must infer her original conclusion and second she must infer something about the causal connection between the conclusion and her evidence. Her second conclusion is her “reconstruction” of the causal connection. But how detailed must her reconstruction be? If she must reconstruct every detail of the causal connection between evidence and conclusion, she will never gain knowledge by way of inference. If she need only reconstruct some “causal connection,” she will always know, since she will always be able to infer that evidence and conclusion are both entailed by their conjunction.

I suggest that it is a mistake to approach the problem as a problem about what else Mary needs to infer before she has knowledge of her original conclusion. Goldman’s remark about reconstructing the causal connection makes more sense as a remark about the kind of inference Mary needs to reach her original conclusion in the first place. It has something to do with principle *P* and the natural account of the Gettier examples.

Nogot presents Mary with evidence that he owns a Ford. She infers that one of her friends owns a Ford. She is justified in reaching that conclusion and it is true. However, since it is true, not because Nogot owns a Ford, but because Havit does, Mary fails to come to know that one of her friends owns a Ford. The natural explanation is that she must infer that Nogot owns a Ford and does not know her final conclusion unless her intermediate conclusion is true. According to this natural explanation, Mary’s inference essentially involves the conclusion that Nogot owns a Ford. According to Goldman, her inference essentially involves a conclusion concerning a causal connection. In order to put these ideas together, we must turn Goldman’s theory of knowledge into a theory of inference.

As a first approximation, let us take his remarks about causal connections literally, forgetting for the moment that they include logical connections.

Then let us transmute his causal theory of knowing into the theory that inductive conclusions always take the form $X \text{ causes } Y$, where further conclusions are reached by additional steps of inductive or deductive reasoning. In particular, we may deduce either X or Y from $X \text{ causes } Y$.

This causal theory of inferring provides the following account of why knowledge requires that we be right about an appropriate causal connection. A person knows by inference only if all conclusions essential to that inference are true. That is, his inference must satisfy principle *P*. Since he can legitimately infer his conclusion only if he can first infer certain causal statements, he can know only if he is right about the causal connection expressed by those statements. First, Mary infers that her evidence is a causal result of Nogot’s past ownership of the Ford. From that she deduces that Nogot has owned a Ford. Then she infers that his past ownership has been causally responsible for present ownership; and she deduces that Nogot owns a Ford. Finally, she deduces that one of her friends owns a Ford. She fails to know because she is wrong when she infers that Nogot’s past ownership is responsible for Nogot’s present ownership.

Inference to the best explanatory statement

A better account of inference emerges if we replace “caus_g” with “because.” On the revised account, we infer not just statements of the form $X \text{ causes } Y$ but, more generally, statements of the form $Y \text{ because } X$ or $X \text{ explains } Y$. Inductive inference is conceived as inference to the best of competing explanatory statements. Inference to a causal explanation is a special case.

The revised account squares better with ordinary usage. Nogot’s past ownership helps to explain Mary’s evidence, but it would sound odd to say that it caused that evidence. Similarly, the detective infers that activities of the butler explain these footprints; does he infer that those activities caused the footprints? A scientist explains the properties of water by means of a hypothesis about unobservable particles that make up the water, but it does not seem right to say that facts about those particles cause the properties of water. An observer infers that certain mental states best explain someone’s behavior; but such explanation by reasons might not be causal explanation.

Furthermore, the switch from “cause” to “because” avoids Goldman’s *ad hoc* treatment of knowledge of generalizations. Although there is no causal relation between a generalization and those observed instances which provide us with evidence for the generalization, there is an obvious explanatory relationship. That all emeralds are green does not cause a particular emerald to be green; but it can explain why that emerald is green. And, other things being equal, we can infer a generalization only if it provides the most plausible way to explain our evidence.

We often infer generalizations that explain but do not logically entail their instances, since they are of the form, *In circumstances C, X’s tend to be Y’s*. Such generalizations may be inferred if they provide a sufficiently plausible account of observed instances all things considered. For example, from the fact that doctors have generally been right in the past when they have said that someone is going to get measles, I infer that doctors can normally tell from certain symptoms that someone is going to get measles. More precisely, I infer that doctors have generally been right in the past because they can normally tell from certain symptoms that someone is going to get measles. This is a very weak explanation, but it is a genuine one. Compare it with the pseudo-explanation, “Doctors are generally right when they say someone has measles because they can normally tell from certain symptoms that someone is going to get measles.”

Similarly, I infer that a substance is soluble in water from the fact that it dissolves when I stirred it into some water. That is a real explanation, to be distinguished from the pseudo-explanation, “That substance dissolves in water because it is soluble in water.” Here too a generalization explains an instance without entailing that instance, since water-soluble substances do not always dissolve in water.

Although we cannot simply deduce instances from this sort of generalization, we can often infer that the generalization will explain some new instance. The inference is warranted if the explanatory claim that *X’s tend to be Y’s will explain why the next X will be Y* is sufficiently more plausible than competitors such as *interfering factor Q will prevent the next X from being a Y*. For example, the doctor says that you will get measles. Because doctors are normally right about that sort of thing, I infer that you will. More precisely, I infer that doctors’ normally being able to tell when someone will get measles will explain the doctor’s being right in this case. The competing

explanatory statements here are not other explanations of the doctor’s being right but rather explanations of his being wrong – e.g., because he has misperceived the symptoms, or because you have faked the symptoms of measles, or because these symptoms are the result of some other disease, etc. Similarly, I infer that this sugar will dissolve in my tea. That is, I infer that the solubility of sugar in tea will explain this sugar’s dissolving in the present case. Competing explanations would explain the sugar’s not dissolving – e.g., because there is already a saturated sugar solution there, because the tea is ice-cold, etc.

*Further examples*⁶

I infer that when I scratch this match it will light. My evidence is that this is a Sure-Fire brand match, and in the past Sure-Fire matches have always lit when scratched. However, unbeknownst to me, this particular match is defective. It will not light unless its surface temperature can be raised to six hundred degrees, which is more than can be attained by scratching. Fortunately, as I scratch the match, a burst of Q-radiation (from the sun) strikes the tip, raising surface temperature to six hundred degrees and igniting the match. Did I know that the match would light? Presumably I did not know. I had justified true belief, but not knowledge. On the present account, the explanation of my failure to know is this: I infer that the match will light in the next instance because Sure-Fire matches generally light when scratched. I am wrong about that; that is not why the match will light this time. Therefore, I do not know that it will light.

It is important that our justification can appeal to a simple generalization even when we have false views about the explanation of that generalization. Consider the man who thinks that barometers fall before a rainstorm because of an increase in the force of gravity. He thinks the gravity pulls the mercury down the tube and then, when the force is great enough, pulls rain out of the sky. Although he is wrong about this explanation, the man in question can come to know that it is going to rain when he sees the barometer falling in a particular case. That a man’s belief is based on an inference that cannot give him knowledge (because it infers a false explanation) does not mean that it is not also based on an inference that does give him knowledge (because it infers a true explanation). The man in question has knowledge because he infers

not only the stronger explanation involving gravity but also the weaker explanation. He infers that the explanation of the past correlation between falling barometer and rain is that the falling barometer is normally associated with rain. Then he infers that this weak generalization will be what will explain the correlation between the falling barometer and rain in the next instance.

Notice that if the man is wrong about that last point, because the barometer is broken and is leaking mercury, so that it is just a coincidence that rain is correlated with the falling barometer in the next instance, he does not come to know that it is going to rain.

Another example is the mad-fiend case. Omar falls down drunk in the street. An hour later he suffers a fatal heart attack not connected with his recent drinking. After another hour a mad fiend comes down the street, spies Omar lying in the gutter, cuts off his head, and runs away. Some time later still, you walk down the street, see Omar lying there, and observe that his head has been cut off. You infer that Omar is dead; and in this way you come to know that he is dead. Now there is no causal connection between Omar's being dead and his head's having been cut off. The fact that Omar is dead is not causally responsible for his head's having been cut off, since if he had not suffered that fatal heart attack he still would have been lying there drunk when the mad fiend came along. And having his head cut off did not cause Omar's death, since he was already dead. Nor is there a straightforward logical connection between Omar's being dead and his having his head cut off. (Given the right sorts of tubes, one might survive decapitation.) So it is doubtful that Goldman's causal theory of knowing can account for your knowledge that Omar is dead.

If inductive inference is inference to the best explanatory statement, your inference might be parsed as follows: "Normally, if someone's head is cut off, that person is dead. This generalization accounts for the fact that Omar's having his head cut off is correlated here with Omar's being dead." Relevant competing explanatory statements in this case would not be competing explanations of Omar's being dead. Instead they would seek to explain Omar's not being dead despite his head's having been cut off. One possibility would be that doctors have carefully connected head and body with special tubes so that blood and air get from body to head and back again. You rule out that hypothesis on grounds of explanatory complica-

tions: too many questions left unanswered (why can't you see the tubes? why wasn't it done in the hospital? etc.). If you cannot rule such possibilities out, then you cannot come to know that Omar is dead. And if you do rule them out but they turn out to be true, again you do not come to know. For example, if it is all an elaborate psychological philosophical experiment, which however fails, then you do not come to know that Omar is dead even though he is dead.

Statistical inference

Statistical inference, and knowledge obtained from it, is also better explicated by way of the notion of statistical explanation than by way of the notion of cause or logical entailment. A person may infer that a particular coin is biased because that provides the best statistical explanation of the observed fraction of heads. His conclusion explains his evidence but neither causes nor entails it.

The relevant kind of statistical explanation does not always make what it explains very probable. For example, suppose that I want to know whether I have the fair coin or the weighted coin. It is equally likely that I have either; the probability of getting heads on a toss of the fair coin is $1/2$; and the probability of getting heads on a toss of the weighted coin is $6/10$. I toss the coin 10,000 times. It comes up heads 4,983 times and tails 5,017. I correctly conclude that the coin is the fair one. You would ordinarily think that I could in this way come to know that I have the fair coin. On the theory of inference we have adopted, I infer the best explanation of the observed distribution of heads and tails. But the explanation, that these were random tosses of a fair coin, does not make it probable that the coin comes up heads exactly 4,983 times and tails exactly 5,017 times in 10,000 tosses. The probability of this happening with a fair coin is very small. If we want to accept the idea that inference is inference to the best explanatory statement, we must agree that statistical explanation can cite an explanation that makes what it explains less probable than it makes its denial. In the present case, I do not explain why 4,983 heads have come up rather than some other number of heads. Instead I explain how it happened that 4,983 heads came up, what led to this happening. I do not explain why this happened rather than something else, since the same thing could easily have led to something else.

To return to an example I have used elsewhere, you walk into a casino and see the roulette wheel stop at red fifty times in a row. The explanation may be that the wheel is fixed. It may also be that the wheel is fair and this is one of those times when fifty reds come up on a fair wheel. Given a fair wheel we may expect that to happen sometimes (but not very often). But if the explanation is that the wheel is fair and that this is just one of those times, it says what the sequence of reds is the result of, the "outcome" of. It does not say why fifty reds in a row occurred this time rather than some other time, nor why that particular series occurred rather than any of the $2^{50}-1$ other possible series.

This kind of statistical explanation explains something as the outcome of a chance set-up. The statistical probability of getting the explained outcome is irrelevant to whether or not we explain that outcome, since this kind of explanation is essentially pure nondeterministic explanation. All that is relevant is that the outcome to be explained is one possible outcome given that chance set-up. That is not to say that the statistical probability of an outcome is irrelevant to the explanation of that outcome. It is relevant in this sense: the greater the statistical probability an observed outcome has in a particular chance set-up, the better that set-up explains that outcome.

The point is less a point about statistical explanation than a point about statistical inference. I wish to infer the best of competing statistical explanations of the observed distribution of heads. This observed outcome has different statistical probabilities in the two hypothetical chance set-ups, fair coin or weighted coin. The higher this statistical probability, the better, from the point of view of inference (other things being equal). The statistical probability of an outcome in a particular hypothetical chance set-up is relevant to how good an explanation that chance set-up provides. Here a better explanation is one that is more likely to be inferable. For example, I infer that I have the fair coin. The statistical probability of 4,983 heads on 10,000 tosses of a fair coin is much greater than the statistical probability of that number of heads on 10,000 tosses of the weighted coin. From the point of view of statistical probability, the hypothesis that the coin is fair offers a better explanation of the observed distribution than the hypothesis that the coin is biased. So statistical probability is relevant to statistical explanation. Not that there is no explanation unless

statistical probability is greater than $1/2$. Rather that statistical probability provides a measure of the inferability of a statistical explanation.

According to probability theory, if initially the coin is just as likely to be the fair one or the weighted one and the statistical probability of the observed outcome is much greater for the fair coin than for the weighted coin, the probability that the coin is fair, given the observed evidence, will be very high. We might conclude that the statistical probability of the observed outcome given the fair or weighted coin is only indirectly relevant to my inference, relevant only because of the theoretical connections between those statistical probabilities and the evidential probabilities of the two hypotheses about the coin, given the observed evidence. But that would be to get things exactly backward. No doubt there is a connection between high evidential probability and inference; but, as we have seen, it is not because there is a purely probabilistic rule of acceptance. High probability by itself does not warrant inference. Only explanatory considerations can do that; and the probability relevant to explanation is statistical probability, the probability that is involved in statistical explanation. It is the statistical probabilities of the observed outcome, given the fair and weighted coins, that is directly relevant to inference. The evidential probabilities of the two hypotheses are only indirectly relevant in that they in some sense reflect the inferability of the hypotheses, where that is determined directly by considerations of statistical probability.

Suppose that at first you do not know which of the two coins I have selected. I toss it 10,000 times, getting 4,983 heads and 5,017 tails. You infer that I have the fair coin, and you are right. But the reason for the 4,983 heads is that I am very good at tossing coins to come up whichever way I desire and I deliberately tossed the coin so as to get roughly half heads and half tails. So, even though you have justified true belief, you do not know that I have the fair coin.

If statistical inference were merely a matter of inferring something that has a high probability on the evidence, there would be no way to account for this sort of Gettier example. And if we are to appeal to principle *P*, it must be a conclusion essential to your inference that the observed outcome is the result of a chance set-up involving the fair coin in such a way that the probability of heads is $1/2$. Given a purely probabilistic rule, that conclusion could not be essential, for reasons similar

to those that have already been discussed concerning the Nogot–Havit case. On the other hand, if statistical inference is inference to the best explanation and there is such a thing as statistical explanation even where the statistical probability of what is explained is quite low, then your conclusion about the reason for my getting 4,983 heads is seen to be essential to your inference. Since your explanation of the observed outcome is false, principle *P* accounts for the fact that you do not come to know that the coin is the fair coin even though you have justified true belief.

Conclusion

We are led to construe induction as inference to the best explanation, or more precisely as inference to the best of competing explanatory statements. The conclusion of any single step of such inference is always of the form *Y because X* (or *X explains Y*), from which we may deduce either *X* or *Y*. Inductive reasoning is seen to consist in a sequence of such explanatory conclusions.

We have been led to this conception of induction in an attempt to account for Gettier examples that show something wrong with the idea that knowledge is justified true belief. We have tried to find principles of inference which, together with principle *P*, would explain Gettier's deviant cases. Purely probabilistic rules were easily seen to be inadequate. Goldman's causal theory of knowing, which promised answers to some of Gettier's questions, suggested a causal theory of induction: inductive inference as inference to the best of competing causal statements. Our present version is simply a modification of that, with *explanatory* replacing *causal*. Its strength lies in the fact that it accounts for a variety of inferences, including inferences that involve weak generalizations or statistical hypotheses, in a way that explains Gettier examples by means of principle *P*.

Evidence One Does Not Possess

Three examples

Example (1)

While I am watching him, Tom takes a library book from the shelf and conceals it beneath his coat. Since I am the library detective, I follow him as he walks brazenly past the guard at the front

door. Outside I see him take out the book and smile. As I approach he notices me and suddenly runs away. But I am sure that it was Tom, for I know him well. I saw Tom steal a book from the library and that is the testimony I give before the University Judicial Council. After testifying, I leave the hearing room and return to my post in the library. Later that day, Tom's mother testifies that Tom has an identical twin, Buck. Tom, she says, was thousands of miles away at the time of the theft. She hopes that Buck did not do it; but she admits that he has a bad character.

Do I know that Tom stole the book? Let us suppose that I am right. It was Tom that took the book. His mother was lying when she said that Tom was thousands of miles away. I do not know that she was lying, of course, since I do not know anything about her, even that she exists. Nor does anyone at the hearing know that she is lying, although some may suspect that she is. In these circumstances I do not know that Tom stole the book. My knowledge is undermined by evidence I do not possess.⁷

Example (2)

Donald has gone off to Italy. He told you ahead of time that he was going; and you saw him off at the airport. He said he was to stay for the entire summer. That was in June. It is now July. Then you might know that he is in Italy. It is the sort of thing one often claims to know. However, for reasons of his own Donald wants you to believe that he is not in Italy but in California. He writes several letters saying that he has gone to San Francisco and has decided to stay there for the summer. He wants you to think that these letters were written by him in San Francisco, so he sends them to someone he knows there and has that person mail them to you with a San Francisco postmark, one at a time. You have been out of town for a couple of days and have not read any of the letters. You are now standing before the pile of mail that arrived while you were away. Two of the phony letters are in the pile. You are about to open your mail. I ask you, "Do you know where Donald is?" "Yes," you reply, "I know that he is in Italy." You are right about where Donald is and it would seem that your justification for believing that Donald is in Italy makes no reference to letters from San Francisco. But you do not know that Donald is in Italy. Your knowledge is undermined by evidence you do not as yet possess.

Example (3)

A political leader is assassinated. His associates, fearing a coup, decide to pretend that the bullet hit someone else. On nationwide television they announce that an assassination attempt has failed to kill the leader but has killed a secret service man by mistake. However, before the announcement is made, an enterprising reporter on the scene telephones the real story to his newspaper, which has included the story in its final edition. Jill buys a copy of that paper and reads the story of the assassination. What she reads is true and so are her assumptions about how the story came to be in the paper. The reporter, whose by-line appears, saw the assassination and dictated his report, which is now printed just as he dictated it. Jill has justified true belief and, it would seem, all her intermediate conclusions are true. But she does not know that the political leader has been assassinated. For everyone else has heard about the televised announcement. They may also have seen the story in the paper and, perhaps, do not know what to believe; and it is highly implausible that Jill should know simply because she lacks evidence everyone else has. Jill does not know. Her knowledge is undermined by evidence she does not possess.

These examples pose a problem for my strategy. They are Gettier examples and my strategy is to make assumptions about inference that will account for Gettier examples by means of principle *P*. But these particular examples appear to bring in considerations that have nothing to do with conclusions essential to the inference on which belief is based.

Some readers may have trouble evaluating these examples. Like other Gettier examples, these require attention to subtle facts about ordinary usage; it is easy to miss subtle differences if, as in the present instance, it is very difficult to formulate a theory that would account for these differences. We must compare what it would be natural to say about these cases if there were no additional evidence one does not possess (no testimony from Tom's mother, no letters from San Francisco, and no televised announcement) with what it would be natural to say about the cases in which there is the additional evidence one does not possess. We must take care not to adopt a very skeptical attitude nor become too lenient about what is to count as knowledge. If we become skeptically inclined, we

will deny there is knowledge in either case. If we become too lenient, we will allow that there is knowledge in both cases. It is tempting to go in one or the other of these directions, toward skepticism or leniency, because it proves so difficult to see what general principles are involved that would mark the difference. But at least some difference between the cases is revealed by the fact that we are *more inclined* to say that there is knowledge in the examples where there is no undermining evidence a person does not possess than in the examples where there is such evidence. The problem, then, is to account for this difference in our inclination to ascribe knowledge to someone.

Evidence against what one knows

If I had known about Tom's mother's testimony, I would not have been justified in thinking that it was Tom I saw steal the book. Once you read the letters from Donald in which he says he is in San Francisco, you are no longer justified in thinking that he is in Italy. If Jill knew about the television announcement, she would not be justified in believing that the political leader has been assassinated. This suggests that we can account for the preceding examples by means of the following principle.

One knows only if there is no evidence such that if one knew about the evidence one would not be justified in believing one's conclusion.

However, by modifying the three examples it can be shown that this principle is too strong.

Suppose that Tom's mother was known to the Judicial Council as a pathological liar. Everyone at the hearing realizes that Buck, Tom's supposed twin, is a figment of her imagination. When she testifies no one believes her. Back at my post in the library, I still know nothing of Tom's mother or her testimony. In such a case, my knowledge would not be undermined by her testimony; but if I were told only that she had just testified that Tom has a twin brother and was himself thousands of miles away from the scene of the crime at the time the book was stolen, I would no longer be justified in believing as I now do that Tom stole the book. Here I know even though there is evidence which, if I knew about it, would cause me not to be justified in believing my conclusion.

Suppose that Donald had changed his mind and never mailed the letters to San Francisco. Then those letters no longer undermine your knowledge. But it is very difficult to see what principle accounts for this fact. How can letters in the pile on the table in front of you undermine your knowledge while the same letters in a pile in front of Donald do not? If you knew that Donald had written letters to you saying that he was in San Francisco, you would not be justified in believing that he was still in Italy. But that fact by itself does not undermine your present knowledge that he is in Italy.

Suppose that as the political leader's associates are about to make their announcement, a saboteur cuts the wire leading to the television transmitter. The announcement is therefore heard only by those in the studio, all of whom are parties to the deception. Jill reads the real story in the newspaper as before. Now, she does come to know that the political leader has been assassinated. But if she had known that it had been announced that he was not assassinated, she would not have been justified in believing that he was, simply on the basis of the newspaper story. Here, a cut wire makes the difference between evidence that undermines knowledge and evidence that does not undermine knowledge.

We can know that h even though there is evidence e that we do not know about such that, if we did know about e , we would not be justified in believing h . If we know that h , it does not follow that we know that there is not any evidence like e . This can seem paradoxical, for it can seem obvious that, if we know that h , we know that any evidence against h can only be misleading. So, later if we get that evidence we ought to be able to know enough to disregard it.

A more explicit version of this interesting paradox goes like this.⁸ "If I know that h is true, I know that any evidence against h is evidence against something that is true; so I know that such evidence is misleading. But I should disregard evidence that I know is misleading. So, once I know that h is true, I am in a position to disregard any future evidence that seems to tell against h ." This is paradoxical, because I am never in a position simply to disregard any future evidence even though I do know a great many different things.

A skeptic might appeal to this paradox in order to argue that, since we are never in a position to disregard any further evidence, we never know anything. Some philosophers would turn the argu-

ment around to say that, since we often know things, we are often in a position to disregard further evidence. But both of these responses go wrong in accepting the paradoxical argument in the first place.

I can know that Tom stole a book from the library without being able automatically to disregard evidence to the contrary. You can know that Donald is in Italy without having the right to ignore whatever further evidence may turn up. Jill may know that the political leader has been assassinated even though she would cease to know this if told that there was an announcement that only a secret service agent had been shot.

The argument for paradox overlooks the way actually having evidence can make a difference. Since I now know that Tom stole the book, I now know that any evidence that appears to indicate something else is misleading. That does not warrant me in simply disregarding any further evidence, since getting that further evidence can change what I know. In particular, after I get such further evidence I may no longer know that it is misleading. For having the new evidence can make it true that I no longer know that Tom stole the book; if I no longer know that, I no longer know that the new evidence is misleading.

Therefore, we cannot account for the problems posed by evidence one does not possess by appeal to the principle, which I now repeat:

One knows only if there is no evidence such that if one knew about the evidence one would not be justified in believing one's conclusion.

For one can know even though such evidence exists.

A result concerning inference

When does evidence one doesn't have keep one from having knowledge? I have described three cases, each in two versions, in which there is misleading evidence one does not possess. In the first version of each case the misleading evidence undermines someone's knowledge. In the second version it does not. What makes the difference?

My strategy is to account for Gettier examples by means of principle P . This strategy has led us to conceive of induction as inference to the best explanation. But that conception of inference

does not by itself seem able to explain these examples. So I want to use the examples in order to learn something more about inference, in particular about what other conclusions are essential to the inference that Tom stole the book, that Donald is in Italy, or that the political leader has been assassinated.

It is not plausible that the relevant inferences should contain essential intermediate conclusions that refer explicitly to Tom's mother, to letters from San Francisco, or to special television programs. For it is very likely that there is an infinite number of ways a particular inference might be undermined by misleading evidence one does not possess. If there must be a separate essential conclusion ruling out each of these ways, inferences would have to be infinitely inclusive – and that is implausible.

Therefore it would seem that the relevant inferences must rule out undermining evidence one does not possess by means of a single conclusion, essential to the inference, that characterizes all such evidence. But how might this be done? It is not at all clear what distinguishes evidence that undermines knowledge from evidence that does not. How is my inference to involve an essential conclusion that rules out Tom's mother's testifying a certain way before a believing audience but does not rule out (simply) her testifying in that way? Or that rules out the existence of letters of a particular sort in the mail on your table but not simply the existence of those letters? Or that rules out a widely heard announcement of a certain sort without simply ruling out the announcement?

Since I am unable to formulate criteria that would distinguish among these cases, I will simply label cases of the first kind "undermining evidence one does not possess." Then we can say this: one knows only if there is no undermining evidence one does not possess. If there is such evidence, one does not know. However, these remarks are completely trivial.

It is somewhat less trivial to use the same label to formulate a principle concerned with inference.

Q One may infer a conclusion only if one also infers that there is no undermining evidence one does not possess.

There is of course an obscurity in principle *Q*; but the principle is not as trivial as the remarks of the last paragraph, since the label "undermining evidence one does not possess" has been explained in

terms of knowledge, whereas this is a principle concerning inference.

If we take principle *Q*, concerning inference, to be basic, we can use principle *P* to account for the differences between the two versions of each of the three examples described above. In each case an inference involves essentially the claim that there is no undermining evidence one does not possess. Since this claim is false in the first version of each case and true in the second, principle *P* implies that there can be knowledge only in the second version of each case.

So there is, according to my strategy, some reason to think that there is a principle concerning inference like principle *Q*. That raises the question of whether there is any independent reason to accept such a principle; and reflection on good scientific practice suggests a positive answer. It is a commonplace that a scientist should base his conclusions on all the evidence. Furthermore, he should not rest content with the evidence he happens to have but should try to make sure he is not overlooking any relevant evidence. A good scientist will not accept a conclusion unless he has some reason to think that there is no as yet undiscovered evidence which would undermine his conclusion. Otherwise he would not be warranted in making his inference. So good scientific practice reflects the acceptance of something like principle *Q*, which is the independent confirmation we wanted for the existence of this principle.

Notice that the scientist must accept something like principle *Q*, with its reference to "undermining evidence one does not possess." For example, he cannot accept the following principle,

One may infer a conclusion only if one also infers that there is no evidence at all such that if he knew that evidence he could not accept his conclusion.

There will always be a true proposition such that if he learned that the proposition was true (and learned nothing else) he would not be warranted in accepting his conclusion. If *h* is his conclusion, and if *k* is a true proposition saying what ticket will win the grand prize in the next New Jersey State Lottery, then *either k or not h* is such a proposition. If he were to learn that it is true that *either k or not h* (and learned nothing else), *not h* would become probable since (given what he knows) *k* is antecedently very improbable. So he could no longer reasonably infer that *h* is true.

There must be a certain kind of evidence such that the scientist infers there is no as yet undiscovered evidence of that kind against *h*. Principle *Q* says that the relevant kind is what I have been labelling "undermining evidence one does not possess." Principle *Q* is confirmed by the fact that good scientific practice involves some such principle and by the fact that principle *Q* together with principle *P* accounts for the three Gettier examples I have been discussing.

If this account in terms of principles *P* and *Q* is accepted, inductive conclusions must involve some self-reference. Otherwise there would be a regress. Before we could infer that *h*, we would have to infer that there is no undermining evidence to *h*. That prior inference could not be deductive, so it would have to be inference to the best explanatory statement. For example, we might infer that the fact that there is no sign of undermining evidence we do not possess is explained by there not being any such evidence. But, then, before we could

accept that conclusion we would first have to infer that there is no undermining evidence to *it* which one does not possess. And, since that inference would have to be inference to the best explanation, it would require a previous inference that there is no undermining evidence for its conclusion; and so on *ad infinitum*.

Clearly, we do not *first* have to infer that there is no undermining evidence to *h* and only then infer *h*. For that would automatically yield the regress. Instead, we must at the same time infer both *h* and that there is no undermining evidence. Furthermore, we infer that there is not only no undermining evidence to *h* but also no undermining evidence to the whole conclusion. In other words, all legitimate inductive conclusions take the form of a self-referential conjunction whose first conjunct is *h* and whose second conjunct (usually left implicit) is the claim that there is no undermining evidence to the whole conjunction.

Notes

- 1 Henry Kyburg, *Probability and the Logic of Rational Belief* (Middletown: Wesleyan University Press, 1961).
- 2 E.g., by Robert Nozick.
- 3 Edmond Gettier, "Is Justified True Belief Knowledge?", this vol., ch. 7.
- 4 Keith Lehrer, "Knowledge, Truth, and Evidence," *Analysis* 25 (1965), pp. 168-75.
- 5 Alvin Goldman, "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1967), pp. 357-72.
- 6 Brian Skyrms, "The Explication of 'X knows that *p*,'" *Journal of Philosophy* 64 (1967), pp. 373-89.
- 7 Keith Lehrer and Thomas Paxson, Jr., "Knowledge: Undefeated Justified True Belief," *Journal of Philosophy* 66 (1969).
- 8 Here and in what follows I am indebted to Saul Kripke, who is, however, not responsible for any faults in my presentation.

Knowledge and Skepticism

Robert Nozick

Knowledge

Conditions for knowledge

Our task is to formulate further conditions to go alongside

- (1) p is true
- (2) S believes that p .

We would like each condition to be necessary for knowledge, so any case that fails to satisfy it will not be an instance of knowledge. Furthermore, we would like the conditions to be jointly sufficient for knowledge, so any case that satisfies all of them will be an instance of knowledge. We first shall formulate conditions that seem to handle ordinary cases correctly, classifying as knowledge cases which are knowledge, and as nonknowledge cases which are not; then we shall check to see how these conditions handle some difficult cases discussed in the literature.¹

The causal condition on knowledge, previously mentioned, provides an inhospitable environment for mathematical and ethical knowledge, also there are well-known difficulties in specifying the type of causal connection. If someone floating in a tank oblivious to everything around him is given (by direct electrical and chemical stimulation of the brain) the belief that he is floating in a tank with his brain being stimulated, then even though that

fact is part of the cause of his belief, still he does not know that it is true.

Let us consider a different third condition

- (3) If p weren't true, S wouldn't believe that p

Throughout this work, let us write the subjunctive "if-then" by an arrow, and the negation of a sentence by prefacing "not-" to it. The above condition thus is rewritten as:

- (3) $\text{not-}p \rightarrow \text{not-(S believes that } p\text{)}$.

This subjunctive condition is not unrelated to the causal condition. Often when the fact that p (partially) causes someone to believe that p , the fact also will be causally necessary for his having the belief — without the cause, the effect would not occur. In that case, the subjunctive condition 3 also will be satisfied. Yet this condition is not equivalent to the causal condition. For the causal condition will be satisfied in cases of causal overdetermination, where either two sufficient causes of the effect actually operate, or a back-up cause (of the same effect) would operate if the first one didn't; whereas the subjunctive condition need not hold for these cases.² When the two conditions do agree, causality indicates knowledge because it acts in a manner that makes the subjunctive 3 true.

The subjunctive condition 3 serves to exclude cases of the sort first described by Edward Gettier, such as the following. Two other people are in my office and I am justified on the basis of much evidence in believing the first owns a Ford car; though he (now) does not, the second person (a stranger to me) owns one. I believe truly and justifiably that someone (or other) in my office owns a Ford car, but I do not know someone

Originally published in R. Nozick, *Philosophical Explanations* (Cambridge MA: Harvard University Press, 1981), pp. 172–85, 197–217; reprinted by permission of the publisher, copyright © by the President and Fellows of Harvard College.

does. Concluded Gettier, knowledge is not simply justified true belief.

The following subjunctive, which specifies condition 3 for this Gettier case, is not satisfied: if no one in my office owned a Ford car, I wouldn't believe that someone did. The situation that would obtain if no one in my office owned a Ford is one where the stranger does not (or where he is not in the office); and in that situation I still would believe, as before, that someone in my office does own a Ford, namely, the first person. So the subjunctive condition 3 excludes this Gettier case as a case of knowledge.

The subjunctive condition is powerful and intuitive, not so easy to satisfy, yet not so powerful as to rule out everything as an instance of knowledge. A subjunctive conditional "if p were true, q would be true," $p \rightarrow q$, does not say that p entails q or that it is logically impossible that p yet not- q . It says that in the situation that would obtain if p were true, q also would be true. This point is brought out especially clearly in recent "possible-worlds" accounts of subjunctives: the subjunctive is true when (roughly) in all those worlds in which p holds true that are closest to the actual world, q also is true. (Examine those worlds in which p holds true closest to the actual world, and see if q holds true in all these.) Whether or not q is true in p worlds that are still farther away from the actual world is irrelevant to the truth of the subjunctive. I do not mean to endorse any particular possible-worlds account of subjunctives, nor am I committed to this type of account.³ I sometimes shall use it, though, when it illustrates points in an especially clear way.⁴

The subjunctive condition 3 also handles nicely cases that cause difficulties for the view that you know that p when you can rule out the relevant alternatives to p in the context. For, as Gail Stine writes,

what makes an alternative relevant in one context and not another? . . . if on the basis of visual appearances obtained under optimum conditions while driving through the countryside Henry identifies an object as a barn, normally we say that Henry knows that it is a barn. Let us suppose, however, that unknown to Henry, the region is full of expertly made papier-mâché facsimiles of barns. In that case, we would not say that Henry knows that the object is a barn, unless he has evidence against it being a papier-mâché facsimile, which is now a relevant alter-

native. So much is clear, but what if no such facsimiles exist in Henry's surroundings, although they once did? Are either of these circumstances sufficient to make the hypothesis (that it's a papier-mâché object) relevant? Probably not, but the situation is not so clear.⁵

Let p be the statement that the object in the field is a (real) barn, and q the one that the object in the field is a papier-mâché barn. When papier-mâché barns are scattered through the area, if p were false, q would be true or might be. Since in this case (we are supposing) the person still would believe p , the subjunctive

(3) not- $p \rightarrow$ not-(S believes that p)

is not satisfied, and so he doesn't know that p . However, when papier-mâché barns are or were scattered around another country, even if p were false q wouldn't be true, and so (for all we have been told) the person may well know that p . A hypothesis q contrary to p clearly is relevant when if p weren't true, q would be true; when not- $p \rightarrow q$. It clearly is irrelevant when if p weren't true, q also would not be true; when not $p \rightarrow$ not- q . The remaining possibility is that neither of these opposed subjunctives holds; q might (or might not) be true if p weren't true. In this case, q also will be relevant, according to an account of knowledge incorporating condition 3 and treating subjunctives along the lines sketched above. Thus, condition 3 handles cases that befuddle the "relevant alternatives" account; though that account can adopt the above subjunctive criterion for when an alternative is relevant, it then becomes merely an alternate and longer way of stating condition 3.⁶

Despite the power and intuitive force of the condition that if p weren't true the person would not believe it, this condition does not (in conjunction with the first two conditions) rule out every problem case. There remains, for example, the case of the person in the tank who is brought to believe, by direct electrical and chemical stimulation of his brain, that he is in the tank and is being brought to believe things in this way; he does not know this is true. However, the subjunctive condition is satisfied: if he weren't floating in the tank, he wouldn't believe he was.

The person in the tank does not know he is there, because his belief is not (sensitive to the truth.) Although it is caused by the fact that is its content, it is not sensitive to that fact. The operators of the tank could have produced any belief,

including the false belief that he wasn't in the tank; if they had, he would have believed that. Perfect sensitivity would involve beliefs and facts varying together. We already have one portion of that variation, subjunctively at least: if p were false he wouldn't believe it. This sensitivity as specified by a subjunctive does not have the belief vary with the truth or falsity of p in all possible situations, merely in the ones that would or might obtain if p were false.

The subjunctive condition

- (3) $\text{not-}p \rightarrow \text{not-}(S \text{ believes that } p)$

tells us only half the story about how his belief is sensitive to the truth-value of p . It tells us how his belief state is sensitive to p 's falsity, but not how it is sensitive to p 's truth; it tells us what his belief state would be if p were false, but not what it would be if p were true.

To be sure, conditions 1 and 2 tell us that p is true and he does believe it, but it does not follow that his believing p is sensitive to p 's being true. This additional sensitivity is given to us by a further subjunctive: if p were true, he would believe it.

- (4) $p \rightarrow S \text{ believes that } p.$

Not only is p true and S believes it, but if it were true he would believe it. Compare: not only was the photon emitted and did it go to the left, but (it was then true that): if it were emitted it would go to the left. The truth of antecedent and consequent is not alone sufficient for the truth of a subjunctive; 4 says more than 1 and 2.⁷ Thus, we presuppose some (or another) suitable account of subjunctives. According to the suggestion tentatively made above, 4 holds true if not only does he actually truly believe p , but in the "close" worlds where p is true, he also believes it. He believes that p for some distance out in the p neighborhood of the actual world; similarly, condition 3 speaks not of the whole not- p neighborhood of the actual world, but only of the first portion of it. (If, as is likely, these explanations do not help, please use your own intuitive understanding of the subjunctives 3 and 4.)

The person in the tank does not satisfy the subjunctive condition 4. Imagine as actual a world in which he is in the tank and is stimulated to believe he is, and consider what subjunctives are true in that world. It is not true of him there that if he were in the tank he would believe it; for in the close world (or situation) to his own where he is in

the tank but they don't give him the belief that he is (much less instill the belief that he isn't) he doesn't believe he is in the tank. Of the person actually in the tank and believing it, it is not true to make the further statement that if he were in the tank he would believe it – so he does not know he is in the tank.⁸

The subjunctive condition 4 also handles a case presented by Gilbert Harman.⁹ The dictator of a country is killed; in their first edition, newspapers print the story, but later all the country's newspapers and other media deny the story, falsely. Everyone who encounters the denial believes it (or does not know what to believe and so suspends judgment). Only one person in the country fails to hear any denial and he continues to believe the truth. He satisfies conditions 1 through 3 (and the causal condition about belief) yet we are reluctant to say he knows the truth. The reason is that if he had heard the denials, he too would have believed them, just like everyone else. His belief is not sensitively tuned to the truth, he doesn't satisfy the condition that if it were true he would believe it. Condition 4 is not satisfied.¹⁰

There is a pleasing symmetry about how this account of knowledge relates conditions 3 and 4, and connects them to the first two conditions. The account has the following form.

- (1)
 (2)
 (3) $\text{not-}1 \rightarrow \text{not-}2$
 (4) $1 \rightarrow 2$

I am not inclined, however, to make too much of this symmetry, for I found also that with other conditions experimented with as a possible fourth condition there was some way to construe the resulting third and fourth conditions as symmetrical answers to some symmetrical looking questions, so that they appeared to arise in parallel fashion from similar questions about the components of true belief.

Symmetry, it seems, is a feature of a mode of presentation, not of the contents presented. A uniform transformation of symmetrical statements can leave the results nonsymmetrical. But if symmetry attaches to mode of presentation, how can it possibly be a deep feature of, for instance, laws of nature that they exhibit symmetry? (One of my favorite examples of symmetry is due to Groucho Marx. On his radio program he spoofed a commercial, and ended, "And if you are not completely satisfied, return the unused portion of our

product and we will return the unused portion of your money.”) Still, to present our subject symmetrically makes the connection of knowledge to true belief especially perspicuous. It seems to me that a symmetrical formulation is a sign of our understanding, rather than a mark of truth. If we cannot understand an asymmetry as arising from an underlying symmetry through the operation of a particular factor, we will not understand why that asymmetry exists in that direction. (But do we also need to understand why the underlying asymmetrical factor holds instead of its opposite?)

A person knows that p when he not only does truly believe it, but also would truly believe it and wouldn't falsely believe it. He not only actually has a true belief, he subjunctively has one. It is true that p and he believes it; if it weren't true he wouldn't believe it, and if it were true he would believe it. To know that p is to be someone who would believe it if it were true, and who wouldn't believe it if it were false.

It will be useful to have a term for this situation when a person's belief is thus subjunctively connected to the fact. Let us say of a person who believes that p , which is true, that when 3 and 4 hold, his belief *tracks* the truth that p . To know is to have a belief that tracks the truth. Knowledge is a particular way of being connected to the world, having a specific real factual connection to the world: tracking it.

One refinement is needed in condition 4. It may be possible for someone to have contradictory beliefs, to believe p and also believe not- p . We do not mean such a person to easily satisfy 4, and in any case we want his belief-state, sensitive to the truth of p , to focus upon p . So let us rewrite our fourth condition as:

- (4) $p \rightarrow S$ believes that p and not-(S believes that not- p).¹¹

As you might have expected, this account of knowledge as tracking requires some refinements and epicycles. Readers who find themselves (or me) bogged down in these refinements should move directly to this essay's second part, on skepticism, where the pace picks up.

Ways and methods

The fourth condition says that if p were true the person would believe it. Suppose the person only happened to see a certain event or simply chanced

on a book describing it. He knows it occurred. Yet if he did not happen to glance that way or encounter the book, he would not believe it, even though it occurred. As written, the fourth condition would exclude this case as one where he actually knows the event occurred. It also would exclude the following case. Suppose some person who truly believes that p would or might arrive at a belief about it in some other close situation where it holds true, in a way or by a method different from the one he (actually) used in arriving at his belief that p , and so thereby come to believe that not- p . In that (close) situation, he would believe not- p even though p still holds true. Yet, all this does not show he actually doesn't know that p , for actually he has not used this alternative method in arriving at his belief. Surely he can know that p , even though condition 4, as written, is not satisfied.

Similarly, suppose he believes that p by one method or way of arriving at belief, yet if p were false he wouldn't use this method but would use another one instead, whose application would lead him mistakenly to believe p (even though it is false). This person does not satisfy condition 3 as written; it is not true of him that if p were false he wouldn't believe it. Still, the fact that he would use another method of arriving at belief if p were false does not show he didn't know that p when he used this method. A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset. Yet this does not mean she doesn't know he is well (or at least ambulatory) when she sees him. Clearly, we must restate our conditions to take explicit account of the ways and methods of arriving at belief.

Let us define a technical locution, S knows, via method (or way of believing) M , that p :

- (1) p is true.
- (2) S believes, via method or way of coming to believe M , that p .
- (3) If p weren't true and S were to use M to arrive at a belief whether (or not) p , then S wouldn't believe, via M , that p .
- (4) If p were true and S were to use M to arrive at a belief whether (or not) p , then S would believe, via M , that p .

We need to relate this technical locution to our ordinary notion of knowledge. If only one method M is actually or subjunctively relevant to S 's belief that p , then, simply, S knows that p (according to

our ordinary notion) if and only if that method M is such that S knows that p via M .

Some situations involve multiple methods, however.

First Situation: S 's belief that p is overdetermined; it was introduced (or reinforced) by two methods, each of which in isolation would have been sufficient to produce in S the belief that p . S 's belief that p via one of these methods satisfies conditions 1–4. However, S 's belief that p via the second method does not satisfy conditions 1–4, and in particular violates condition 3.

A case of this sort is discussed by Armstrong.¹² A father believes his son innocent of committing a particular crime, both because of faith in his son and (now) because he has seen presented in the courtroom a conclusive demonstration of his son's innocence. His belief via the method of courtroom demonstration satisfies 1–4, let us suppose, but his faith-based belief does not. If his son were guilty, he would still believe him innocent, on the basis of faith in his son. Thus, his belief that p (that his son is innocent) via faith in his son violates condition 3. Looking at his belief alone, without mention of method, his belief that p violates the third condition (namely, if p were false S wouldn't believe that p), which made no mention of method.

Second Situation: S 's belief that p via one method satisfies conditions 1–4. However, if p were false, S would not use that method in arriving at a belief about the truth value of p . Instead, he would use another method, thereby deciding, despite p 's falsity, that p was true. S 's actual belief that p is in no way based on the use of this second method, but if p were false he would believe p via the second method. (However, if p were false and S were to decide about its truth value by using the first method, then S would not believe that p . To be sure, if p were false S wouldn't decide about it by using that first method.) The truth value of p affects which method S uses to decide whether p .

Our earlier example of the grandmother is of this sort. Consider one further example, suggested to me by Avishai Margalit. S believes a certain building is a theater and concert hall. He has attended plays and concerts there (first method). However, if the building were not a theater, it would have housed a nuclear reactor that would so have altered

the air around it (let us suppose) that everyone upon approaching the theater would have become lethargic and nauseous, and given up the attempt to buy a ticket. The government cover story would have been that the building was a theater, a cover story they knew would be safe since no unmedicated person could approach through the nausea field to discover any differently. Everyone, let us suppose, would have believed the cover story; they would have believed that the building they saw (but only from some distance) was a theater.

S believes the building is a theater because he has attended plays and concerts inside. He does not believe it is a theater via the second method of reading the government's cover story plus planted spurious theater and concert reviews. There are no such things. However, if it weren't a theater, it would be a nuclear reactor, there would be such cover stories, and S would believe still (this time falsely and via the second method) that the building was a theater. Nonetheless, S , who actually has attended performances there, knows that it is a theater.

To hold that a person knows that p if there exists at least one method M , satisfying conditions 1–4, via which he believes that p , would classify the father as knowing his son is innocent, a consequence too charitable to the father. Whereas it seems too stringent to require that all methods satisfy conditions 1–4, including those methods that were not actually used but would be under some other circumstances; the grandmother knows her grandson is well, and the person who has attended the concerts and plays knows the building is a theater. It is more reasonable to hold he knows that p if all the methods via which he actually believes that p satisfy conditions 1–4. Yet suppose our theatergoer also believes it is a theater partly because government officials, before they decided on which use they would put the building to, announced they were building a theater. Still, the theatergoer knows the building is a theater. Not all methods actually used need satisfy conditions 1–4, but we already have seen how the weak position that merely one such method is enough mishandles the case of the father.

We are helped to thread our way through these difficulties when we notice this father does not merely believe his son is innocent via the route of faith in his son; this defective route, not satisfying 1–4, also outweighs for him the method of courtroom demonstration. Even if courtroom demonstration (had it operated alone) would lead to the

Table 10.1

	<i>M</i> ₁ recommends	<i>M</i> ₂ recommends	Does the person believe <i>p</i> or believe not- <i>p</i> ?
Case I	believe <i>p</i>	believe <i>p</i>	believes <i>p</i>
Case II	believe <i>p</i>	believe not- <i>p</i>	?
Case III	believe not- <i>p</i>	believe <i>p</i>	?

belief that his son is guilty, that not-*p*, still he would believe his son innocent, via faith in his son. Although it is the method of courtroom demonstration that gives him knowledge that *p* if anything does, for the father this method is outweighed by faith.¹³ As a first try at delineating outweighing, we might say that method *M* is outweighed by others if when *M* would have the person believe *p*, the person believes not-*p* if the other methods would lead to the belief that not-*p*, or when *M* would have the person believe not-*p*, the person believes *p* if the other methods would lead to the belief that *p*.

This leads us to put forth the following position: S knows that *p* if there is some method via which S believes that *p* which satisfies conditions 1-4, and that method is not outweighed by any other method(s), via which S actually believes that *p*, that fail to satisfy conditions 3 and 4. According to this position, in some cases a person has knowledge even when he also actually believes via a method *M*₁ that does not satisfy 1-4, provided it is outweighed by one that does; namely, in the over-determination case, and in the case when *M*₁ alone would suffice to fix belief but only in the absence of a verdict from the *M* he also uses which does satisfy 1-4.

S knows that *p* if and only if there is a method *M* such that (a) he knows that *p* via *M*, his belief via *M* that *p* satisfies conditions 1-4, and (b) all other methods *M*₁ via which he believes that *p* that do not satisfy conditions 1-4 are outweighed by *M*.¹⁴

We have stated our outweighing requirement only roughly; now we must turn to refinements. According to our rough statement, in the over-determination case, method *M*₁, which satisfies 3 and 4 and which is what gives knowledge if anything does, wins out over the other method *M*₂ in all cases. The actual situation (Case I) is where *M*₁ recommends believing *p* as does *M*₂, and the person believes *p*. In this case we have made our answer to the question whether he knows that *p* depend on what happens or would happen in the two other cases where the methods recommend

different beliefs (see Table 10.1). The first rough statement held that the person knows in Case I only if he would believe *p* in Case II and not-*p* in Case III. While this is sufficient for knowledge in Case I, it seems too stringent to be necessary for such knowledge.

An alternative and more adequate view would hold constant what the other method recommends, and ask whether the belief varies with the recommendation of *M*₁. Since *M*₂ actually recommends *p* (Case I), we need look only at Case III and ask: when *M*₂ continues to recommend *p* and *M*₁ recommends not-*p*, would the person believe not-*p*? Despite his faith, would the father believe his son guilty if the courtroom procedure proved guilt? That is the relevant question – not what he would believe if the courtroom showed innocence while (somehow) his method of faith led to a conclusion of guilty.

Consider how this works out in another simple case. I see a friend today; he is now alive. However, if he were not alive, I wouldn't have seen him today or (let us suppose) heard of his death, and so still would believe he was alive. Yet condition 3 is satisfied; it includes reference to a method, and the method *M*₁ of seeing him satisfies 3 with respect to *p* equals he is alive at the time. But there also is another method *M*₂ via which I believe he is alive, namely having known he was alive yesterday and continuing to believe it. Case III asks what I would believe if I saw the friend dead (though I knew yesterday he was alive); our position holds I must believe him dead in this case if I am to know by seeing him that he is alive in Case I. However, we need not go so far as to consider what I would believe if I had "learned" yesterday that he was dead yet "saw" him alive today. Perhaps in that case I would wonder whether it really was he I was seeing. Even so, given the result in Case III, I know (in Case I) he is alive. Thus, we hold fixed the recommendation of the other method, and only ask whether then the belief varies with the recommendation of method *M*₁.¹⁵

Our test of looking at Case III cannot apply if M_1 is a one-sided method, incapable of recommending belief in not- p ; it either recommends belief in p or yields no recommendation. (Perhaps M_1 detects one of a number of sufficient conditions for p ; not detecting this, M_1 remains silent as to the truth of p .) What are we to say about his knowing if a person's belief is overdetermined or jointly determined by a one-sided method M_1 plus another method M_2 which fails to satisfy condition 3? Should we now look at Case II, where M_1 recommends belief in p and M_2 recommends belief in not- p , and say that believing p in this case is sufficient to show that M_1 outweighs M_2 ? That does not seem unreasonable, but we had better be careful to stipulate that this Case II situation is a sufficient condition for M_1 's outweighing M_2 only when the Case III situation is impossible, for otherwise we face the possibility of divergent results. (For example, he believes p in Case II and in Case III, yet believes not- p when both methods recommend not- p ; here the result in Case II indicates M_1 outweighs M_2 while the result in Case III indicates M_2 outweighs M_1 .) It is Case III that should predominate.

One final remark about method. Suppose a method is good for some types of statements but not others; it satisfies 3 and 4 for the first type but not for the second. However, S believes the method is good for all types of statements and applies it indiscriminately. When he applies it to a statement of the first type which he thereby comes to believe, does he know that it is true? He does, if he satisfies conditions 3 and 4. Hesitation to grant him knowledge stems, I think, from the fact that if p were false and were of the second type, he might well still believe it. Whether or not this undercuts condition 3 for knowledge depends upon the disparity of the two types; the greater the gulf between the types, the more willing we are to say he knows a statement of the type where M works.

In explaining the nature of knowledge by reference to a method or way of believing, we leave large questions open about how to individuate methods, count them, identify which method is at work, and so on. I do not want to underestimate these difficulties, but neither do I want to pursue them here.¹⁶ Still, some clarifying remarks are needed.

A person can use a method (in my sense) without proceeding methodically, and without knowledge or awareness of what method he is using.

Usually, a method will have a final upshot in experience on which the belief is based, such as visual experience, and then (a) no method without this upshot is the same method, and (b) any method experientially the same, the same "from the inside," will count as the same method. Basing our beliefs on experiences, you and I and the person floating in the tank are using, for these purposes, the same method.

Some methods are supervenient on others, for example, "believing what seems to be true to you" or "believing what seems true given the weighting of all other methods." The account of outweighing is not to apply to such supervenient methods, otherwise there always will be such a one that outweighs all the others. There are various gerrymandered (Goodmanesque) methods that would yield the same resulting belief in the actual situation; which method a person actually is using will depend on which general disposition to acquire beliefs (extending to other situations) he actually is exercising.¹⁷

Although sometimes it will be necessary to be explicit about the methods via which someone believes something, often it will cause no confusion to leave out all mention of method. Furthermore, some statements play a central role in our continuing activities, or in our picture of the world or framework wherein we check other statements, for example, "I have two hands," "the world has existed for many years already"; it is misleading to think of our coming to believe them via some delimited method or methods.¹⁸ So nested are these statements in our other beliefs and activities, and so do they nest them, that our belief or acceptance of them is (for almost all purposes) best represented apart from any particular methods. In considering our knowledge of them we may revert to the earlier simpler subjunctives

- (3) not- $p \rightarrow$ not-(S believes that p)
- (4) $p \rightarrow$ S believes that p .

The very centrality of the specific p means that 4 will be satisfied without reference to a specific method or way of believing. In contrast, I know there is a pair of scissors on my desk (in front of me) now; but it is not accurate simply to say that if there were a pair of scissors there, I would believe there was. For what if I weren't looking, or hadn't looked, or were elsewhere now? Reference to the method via which I believe there are scissors on the desk is needed to exclude these possibilities. With the most central statements, however, there

is no similar “what if”; their centrality ensures they will not escape notice.

Skepticism

The skeptic about knowledge argues that we know very little or nothing of what we think we know, or at any rate that this position is no less reasonable than the belief in knowledge. The history of philosophy exhibits a number of different attempts to refute the skeptic: to prove him wrong or show that in arguing against knowledge he presupposes there is some and so refutes himself. Others attempt to show that accepting skepticism is unreasonable, since it is more likely that the skeptic’s extreme conclusion is false than that all of his premisses are true, or simply because reasonableness of belief just means proceeding in an anti-skeptical way. Even when these counterarguments satisfy their inventors, they fail to satisfy others, as is shown by the persistent attempts against skepticism.¹⁹ The continuing felt need to refute skepticism, and the difficulty in doing so, attests to the power of the skeptic’s position, the depth of his worries.

An account of knowledge should illuminate skeptical arguments and show wherein lies their force. If the account leads us to reject these arguments, this had better not happen too easily or too glibly. To think the skeptic overlooks something obvious, to attribute to him a simple mistake or confusion or fallacy, is to refuse to acknowledge the power of his position and the grip it can have upon us. We thereby cheat ourselves of the opportunity to reap his insights and to gain self-knowledge in understanding why his arguments lure us so. Moreover, in fact, we cannot lay the specter of skepticism to rest without first hearing what it shall unfold.

Our goal is not, however, to refute skepticism, to prove it is wrong or even to argue that it is wrong. Our task here is to explain how knowledge is possible, given what the skeptic says that we do accept (for example, that it is logically possible that we are dreaming or are floating in the tank). In doing this, we need not convince the skeptic, and we may introduce explanatory hypotheses that he would reject. What is important for our task of explanation and understanding is that *we* find those hypotheses acceptable or plausible, and that they show us how the existence of knowledge fits together with the logical possibilities the skeptic

points to, so that these are reconciled within our own belief system. These hypotheses are to explain to ourselves how knowledge is possible, not to prove to someone else that knowledge *is* possible.²⁰

Skeptical possibilities

The skeptic often refers to possibilities in which a person would believe something even though it was false: really, the person is cleverly deceived by others, perhaps by an evil demon, or the person is dreaming or he is floating in a tank near Alpha Centauri with his brain being stimulated. In each case, the *p* he believes is false, and he believes it even though it *is* false.

How do these possibilities adduced by the skeptic show that someone does not know that *p*? Suppose that someone is you; how do these possibilities count against your knowing that *p*? One way might be the following. (I shall consider other ways later.) If there is a possible situation where *p* is false yet you believe that *p*, then in that situation you believe that *p* even though it is false. So it appears you do not satisfy condition 3 for knowledge.

(3) If *p* were false, S wouldn’t believe that *p*.

For a situation has been described in which you do believe that *p* even though *p* is false. How then can it also be true that if *p* were false, you wouldn’t believe it? If the skeptic’s possible situation shows that 3 is false, and if 3 is a necessary condition for knowledge, then the skeptic’s possible situation shows that there isn’t knowledge.

So construed, the skeptic’s argument plays on condition 3; it aims to show that condition 3 is not satisfied. The skeptic may seem to be putting forth

R: Even if *p* were false, S still would believe *p*.²¹

This conditional, with the same antecedent as 3 and the contradictory consequent, is incompatible with the truth of 3. If 3 is true, then R is not. However, R is stronger than the skeptic needs in order to show 3 is false. For 3 is false when if *p* were false, S might believe that *p*. This last conditional is weaker than R, and is merely 3’s denial:

T: not-[not-*p* → not-(S believes that *p*)].

Whereas R does not simply deny 3, it asserts an opposing subjunctive of its own. Perhaps the possibility the skeptic adduces is not enough to show that R is true, but it appears at least to establish the

weaker T; since this T denies 3, the skeptic's possibility appears to show that 3 is false.²²

However, the truth of 3 is not incompatible with the existence of a possible situation where the person believes p though it is false. The subjunctive

(3) $\text{not-}p \rightarrow \text{not-}(S \text{ believes } p)$

does not talk of all possible situations in which p is false (in which $\text{not-}p$ is true). It does not say that in all possible situations where $\text{not-}p$ holds, S doesn't believe p . To say there is no possible situation in which $\text{not-}p$ yet S believes p , would be to say that $\text{not-}p$ entails $\text{not-}(S \text{ believes } p)$, or logically implies it. But subjunctive conditionals differ from entailments; the subjunctive 3 is not a statement of entailment. So the existence of a possible situation in which p is false yet S believes p does not show that 3 is false;²³ 3 can be true even though there is a possible situation where $\text{not-}p$ and S believes that p .

What the subjunctive 3 speaks of is the situation that would hold if p were false. Not every possible situation in which p is false is the situation that would hold if p were false. To fall into possible worlds talk, the subjunctive 3 speaks of the $\text{not-}p$ world that is closest to the actual world, or of those $\text{not-}p$ worlds that are closest to the actual world, or more strongly (according to my suggestion) of the $\text{not-}p$ neighborhood of the actual world. And it is of this or these $\text{not-}p$ worlds that it says (in them) S does not believe that p . What happens in yet other more distant $\text{not-}p$ worlds is no concern of the subjunctive 3.

The skeptic's possibilities (let us refer to them as SK), of the person's being deceived by a demon or dreaming or floating in a tank, count against the subjunctive

(3) if p were false then S wouldn't believe that p only if (one of) these possibilities would or might obtain if p were false; only if one of these possibilities is in the $\text{not-}p$ neighborhood of the actual world. Condition 3 says: if p were false, S still would not believe p . And this can hold even though there is some situation SK described by the skeptic in which p is false and S believes p . If p were false S still would not believe p , even though there is a situation SK in which p is false and S does believe p , provided that this situation SK wouldn't obtain if p were false. If the skeptic describes a situation SK which would not hold even if p were false then this situation SK doesn't

show that 3 is false and so does not (in this way at least) undercut knowledge. Condition C acts to rule out skeptical hypotheses.

C: $\text{not-}p \rightarrow \text{SK does not obtain.}$

Any skeptical situation SK which satisfies condition C is ruled out. For a skeptical situation SK to show that we don't know that p , it must fail to satisfy C which excludes it; instead it must be a situation that might obtain if p did not, and so satisfy C's denial:

$\text{not-}(\text{not-}p \rightarrow \text{SK doesn't obtain.})$

Although the skeptic's imagined situations appear to show that 3 is false, they do not; they satisfy condition C and so are excluded.

The skeptic might go on to ask whether we know that his imagined situations SK are excluded by condition C, whether we know that if p were false SK would not obtain. However, typically he asks something stronger: do we know that his imagined situation SK does not actually obtain? Do we know that we are not being deceived by a demon, dreaming, or floating in a tank? And if we do not know this, how can we know that p ? Thus we are led to the second way his imagined situations might show that we do not know that p .

Skeptical results

According to our account of knowledge, S knows that the skeptic's situation SK doesn't hold if and only if

- (1) SK doesn't hold
- (2) S believes that SK doesn't hold
- (3) If SK were to hold, S would not believe that SK doesn't hold
- (4) If SK were not to hold, S would believe it does not.

Let us focus on the third of these conditions. The skeptic has carefully chosen his situations SK so that if they held we (still) would believe they did not. We would believe we weren't dreaming, weren't being deceived, and so on, even if we were. He has chosen situations SK such that if SK were to hold, S would (still) believe that SK doesn't hold – and this is incompatible with the truth of 3.²⁴

Since condition 3 is a necessary condition for knowledge, it follows that we do not know that SK doesn't hold. If it were true that an evil demon was deceiving us, if we were having a particular dream,

if we were floating in a tank with our brains stimulated in a specified way, we would still believe we were not. So, we do not know we're not being deceived by an evil demon, we do not know we're not in that tank, and we do not know we're not having that dream. So says the skeptic, and so says our account. And also so we say – don't we? For how could we know we are not being deceived that way, dreaming that dream? If those things *were* happening to us, everything would seem the same to us. There is no way we can know it is not happening for there is no way we could tell if it were happening; and if it were happening we would believe exactly what we do now – in particular, we still would believe that it was not. For this reason, we feel, and correctly, that we don't know – how could we? – that it is not happening to us. It is a virtue of our account that it yields, and explains, this result.

The skeptic asserts we do not know his possibilities don't obtain, and he is right. Attempts to avoid skepticism by claiming we do know these things are bound to fail. The skeptic's possibilities make us uneasy because, as we deeply realize, we do not know they don't obtain; it is not surprising that attempts to show we do know these things leave us suspicious, strike us even as bad faith.²⁵ Nor has the skeptic merely pointed out something obvious and trivial. It comes as a surprise to realize that we do not know his possibilities don't obtain. It is startling, shocking. For we would have thought, before the skeptic got us to focus on it, that we did know those things, that we did know we were not being deceived by a demon, or dreaming that dream, or stimulated that way in that tank. The skeptic has pointed out that we do not know things we would have confidently said we knew. And if we don't know these things, what can we know? So much for the supposed obviousness of what the skeptic tells us.

Let us say that a situation (or world) is doxically identical for S to the actual situation when if S were in that situation, he would have exactly the beliefs (*doxa*) he actually does have. More generally, two situations are doxically identical for S if and only if he would have exactly the same beliefs in them. It might be merely a curiosity to be told there are nonactual situations doxically identical to the actual one. The skeptic, however, describes worlds doxically identical to the actual world in which almost everything believed is false.²⁶

Such worlds are possible because we know mediately, not directly. This leaves room for a

divergence between our beliefs and the truth. It is as though we possessed only two-dimensional plane projections of three-dimensional objects. Different three-dimensional objects, oriented appropriately, have the same two-dimensional plane projection. Similarly, different situations or worlds will lead to our having the very same beliefs. What is surprising is how very different the doxically identical world can be – different enough for almost everything believed in it to be false. Whether or not the mere fact that knowledge is mediated always makes room for such a very different doxically identical world, it does so in our case, as the skeptic's possibilities show. To be shown this is nontrivial, especially when we recall that we do not know the skeptic's possibility doesn't obtain: we do not know that we are not living in a doxically identical world wherein almost everything we believe is false.²⁷

What more could the skeptic ask for or hope to show? Even readers who sympathized with my desire not to dismiss the skeptic too quickly may feel this has gone too far, that we have not merely acknowledged the force of the skeptic's position but have succumbed to it.

The skeptic maintains that we know almost none of what we think we know. He has shown, much to our initial surprise, that we do not know his (nontrivial) possibility SK doesn't obtain. Thus, he has shown of one thing we thought we knew, that we didn't and don't. To the conclusion that we know almost nothing, it appears but a short step. For if we do not know we are not dreaming or being deceived by a demon or floating in a tank, then how can I know, for example, that I am sitting before a page writing with a pen, and how can you know that you are reading a page of a book?

However, although our account of knowledge agrees with the skeptic in saying that we do not know that not-SK, it places no formidable barriers before my knowing that I am writing on a page with a pen. It is true that I am, I believe I am, if I weren't I wouldn't believe I was, and if I were, I would believe it. (I leave out the reference to method.) Also, it is true that you are reading a page (please, don't stop now!), you believe you are, if you weren't reading a page you wouldn't believe you were, and if you were reading a page you would believe you were. So according to the account, I do know that I am writing on a page with a pen, and you do know that you are reading a page. The account does not lead to any general skepticism.

Yet we must grant that it appears that if the skeptic is right that we don't know we are not dreaming or being deceived or floating in the tank, then it cannot be that I know I am writing with a pen or that you know you are reading a page. So we must scrutinize with special care the skeptic's "short step" to the conclusion that we don't know these things, for either this step cannot be taken or our account of knowledge is incoherent.

Nonclosure

In taking the "short step," the skeptic assumes that if S knows that *p* and he knows that *p* entails *q* then he also knows that *q*. In the terminology of the logicians, the skeptic assumes that knowledge is closed under known logical implication; that the operation of moving from something known to something else known to be entailed by it does not take us outside of the (closed) area of knowledge. He intends, of course, to work things backwards, arguing that since the person does not know that *q*, assuming (at least for the purposes of argument) that he does know that *p* entails *q*, it follows that he does not know that *p*. For if he did know that *p*, he would also know that *q*, which he doesn't.

The details of different skeptical arguments vary in their structure, but each one will assume some variant of the principle that knowledge is closed under known logical implication. If we abbreviate "knowledge that *p*" by "K*p*" and abbreviate "entails" by the fish-hook sign " \prec ," we can write this principle of closure as the subjunctive principle

P: $K(p \prec q) \& Kp \rightarrow Kq$.

If a person were to know that *p* entails *q* and he were to know that *p* then he would know that *q*. The statement that *q* follows by modus ponens from the other two stated as known in the antecedent of the subjunctive principle P; this principle counts on the person to draw the inference to *q*.

You know that your being in a tank on Alpha Centauri entails your not being in place X where you are. (I assume here a limited readership.) And you know also the contrapositive, that your being at place X entails that you are not then in a tank on Alpha Centauri. If you knew you were at X you would know you're not in a tank (of a specified sort) at Alpha Centauri. But you do not know this

last fact (the skeptic has argued and we have agreed) and so (he argues) you don't know the first. Another intuitive way of putting the skeptic's argument is as follows. If you know that two statements are incompatible and you know the first is true then you know the denial of the second. You know that your being at X and your being in a tank on Alpha Centauri are incompatible; so if you knew you were at X you would know you were not in the (specified) tank on Alpha Centauri. Since you do not know the second, you don't know the first.²⁸

No doubt, it is possible to argue over the details of principle P, to point out it is incorrect as it stands. Perhaps, though K*p*, the person does not know that he knows that *p* (that is, not-KK*p*) and so does not draw the inference to *q*. Or perhaps he doesn't draw the inference because not-KK($p \prec q$). Other similar principles face their own difficulties: for example, the principle that $K(p \rightarrow q) \rightarrow (Kp \rightarrow Kq)$ fails if K*p* stops $p \rightarrow q$ from being true, that is, if $Kp \rightarrow \text{not}(p \rightarrow q)$; the principle that $K(p \prec q) \rightarrow K(Kp \rightarrow Kq)$ faces difficulties if K*p* makes the person forget that $(p \prec q)$ and so he fails to draw the inference to *q*. We seem forced to pile K upon K until we reach something like $KK(p \prec q) \& KKp \rightarrow Kq$; this involves strengthening considerably the antecedent of P and so is not useful for the skeptic's argument that *p* is not known. (From a principle altered thus, it would follow at best that it is not known that *p* is known.)

We would be ill-advised, however, to quibble over the details of P. Although these details are difficult to get straight, it will continue to appear that something like P is correct. If S knows that "*p* entails *q*" and he knows that *p* and knows that "*p* and *p* entails *q*" entails *q*" and he does draw the inference to *q* from all this and believes *q* via the process of drawing this inference, then will he not know that *q*? And what is wrong with simplifying this mass of detail by writing merely principle P, provided we apply it only to cases where the mass of detail holds, as it surely does in the skeptical cases under consideration? For example, I do realize that my being in the Van Leer Foundation Building in Jerusalem entails that I am not in a tank on Alpha Centauri; I am capable of drawing inferences now; I do believe I am not in a tank on Alpha Centauri (though not solely via this inference, surely); and so forth. Won't this satisfy the correctly detailed principle, and shouldn't it follow that I know I am not (in that tank) on Alpha

Centauri? The skeptic agrees it should follow; so he concludes from the fact that I don't know I am not floating in the tank on Alpha Centauri that I don't know I am in Jerusalem. Uncovering difficulties in the details of particular formulations of P will not weaken the principle's intuitive appeal; such quibbling will seem at best like a wasp attacking a steamroller, at worst like an effort in bad faith to avoid being pulled along by the skeptic's argument.

Principle P is wrong, however, and not merely in detail. Knowledge is not closed under known logical implication.²⁹ S knows that p when S has a true belief that p , and S wouldn't have a false belief that p (condition 3) and S would have a true belief that p (condition 4). Neither of these latter two conditions is closed under known logical implication.

Let us begin with condition

(3) if p were false, S wouldn't believe that p .

When S knows that p , his belief that p is contingent on the truth of p , contingent in the way the subjunctive condition 3 describes. Now it might be that p entails q (and S knows this), that S's belief that p is subjunctively contingent on the truth of p , that S believes q , yet his belief that q is not subjunctively dependent on the truth of q , in that it (or he) does not satisfy:

(3) if q were false, S wouldn't believe that q .

For 3' talks of what S would believe if q were false, and this may be a very different situation than the one that would hold if p were false, even though p entails q . That you were born in a certain city entails that you were born on earth.³⁰ Yet contemplating what (actually) would be the situation if you were not born in that city is very different from contemplating what situation would hold if you weren't born on earth. Just as those possibilities are very different, so what is believed in them may be very different. When p entails q (and not the other way around) p will be a stronger statement than q , and so not- q (which is the antecedent of 3') will be a stronger statement than not- p (which is the antecedent of 3). There is no reason to assume you will have the same beliefs in these two cases, under these suppositions of differing strengths.

There is no reason to assume the (closest) not- p world and the (closest) not- q world are doxically identical for you, and no reason to assume, even though p entails q , that your beliefs in one of these

worlds would be a (proper) subset of your beliefs in the other.

Consider now the two statements:

p = I am awake and sitting on a chair in Jerusalem;

q = I am not floating in a tank on Alpha Centauri being stimulated by electrochemical means to believe that p .

The first one entails the second: p entails q . Also, I know that p entails q ; and I know that p . If p were false, I would be standing or lying down in the same city, or perhaps sleeping there, or perhaps in a neighboring city or town. If q were false, I would be floating in a tank on Alpha Centauri. Clearly these are very different situations, leading to great differences in what I then would believe. If p were false, if I weren't awake and sitting on a chair in Jerusalem, I would not believe that p . Yet if q were false, if I was floating in a tank on Alpha Centauri, I would believe that q , that I was not in the tank, and indeed, in that case, I would still believe that p . According to our account of knowledge, I know that p yet I do not know that q , even though (I know) p entails q .

This failure of knowledge to be closed under known logical implication stems from the fact that condition 3 is not closed under known logical implication; condition 3 can hold of one statement believed while not of another known to be entailed by the first.³¹ It is clear that any account that includes as a necessary condition for knowledge the subjunctive condition 3, not- $p \rightarrow$ not-(S believes that p), will have the consequence that knowledge is not closed under known logical implication.³²

When p entails q and you believe each of them, if you do not have a false belief that p (since p is true) then you do not have a false belief that q . However, if you are to know something not only don't you have a false belief about it, but also you wouldn't have a false belief about it. Yet, we have seen how it may be that p entails q and you believe each and you wouldn't have a false belief that p yet you might have a false belief that q (that is, it is not the case that you wouldn't have one). Knowledge is not closed under the known logical implication because "wouldn't have a false belief that" is not closed under known logical implication.

If knowledge were the same as (simply) true belief then it would be closed under known logical implication (provided the implied statements were

believed). Knowledge is not simply true belief, however; additional conditions are needed. These further conditions will—make knowledge open under known logical implication, even when the entailed statement is believed, when at least one of the further conditions itself is open. Knowledge stays closed (only) if all of the additional conditions are closed. I lack a general nontrivial characterization of those conditions that are closed under known logical implication; possessing such an illuminating characterization, one might attempt to prove that no additional conditions of that sort could provide an adequate analysis of knowledge.

Still, we can say the following. A belief that p is knowledge that p only if it somehow varies with the truth of p . The causal condition for knowledge specified that the belief was “produced by” the fact, but that condition did not provide the right sort of varying with the fact. The subjunctive conditions 3 and 4 are our attempt to specify that varying. But however an account spells this out, it will hold that whether a belief that p is knowledge partly depends on what goes on with the belief in some situations when p is false. An account that says nothing about what is believed in any situation when p is false cannot give us any mode of varying with the fact.

Because what is preserved under logical implication is truth, any condition that is preserved under known logical implication is most likely to speak only of what happens when p , and q , are true, without speaking at all of what happens when either one is false. Such a condition is incapable of providing “varies with”; so adding only such conditions to true belief cannot yield an adequate account of knowledge.³³

A belief’s somehow varying with the truth of what is believed is not closed under known logical implication. Since knowledge that p involves such variation, knowledge also is not closed under known logical implication. The skeptic cannot easily deny that knowledge involves such variation, for his argument that we don’t know that we’re not floating in that tank, for example, uses the fact that knowledge does involve variation. (“If you were floating in the tank you would still think you weren’t, so you don’t know that you’re not.”) Yet, though one part of his argument uses that fact that knowledge involves such variation, another part of his argument presupposes that knowledge does not involve any such variation. This latter is the part that depends upon know-

ledge being closed under known logical implication, as when the skeptic argues that since you don’t know that not-SK, you don’t know you are not floating in the tank, then you also don’t know, for example, that you are now reading a book. That closure can hold only if the variation does not. The skeptic cannot be right both times. According to our view he is right when he holds that knowledge involves such variation and so concludes that we don’t know, for example, that we are not floating in that tank; but he is wrong when he assumes knowledge is closed under known logical implication and concludes that we know hardly anything.³⁴

Knowledge is a real factual relation, subjunctively specifiable, whose structure admits our standing in this relation, tracking, to p without standing in it to some q which we know p to entail. Any relation embodying some variation of belief with the fact, with the truth (value), will exhibit this structural feature. The skeptic is right that we don’t track some particular truths – the ones stating that his skeptical possibilities SK don’t hold – but wrong that we don’t stand in the real knowledge-relation of tracking to many other truths, including ones that entail these first mentioned truths we believe but don’t know.

The literature on skepticism contains writers who endorse these skeptical arguments (or similar narrower ones), but confess their inability to maintain their skeptical beliefs at times when they are not focusing explicitly on the reasoning that led them to skeptical conclusions. The most notable example of this is Hume:

I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another... Most fortunately it happens that since reason is incapable of dispelling these clouds, nature herself suffices to that purpose, and cures me of this philosophical melancholy and delirium, either by relaxing this bent of mind, or by some avocation, and lively impression of my senses, which obliterate all these chimeras. I dine, I play a game of backgammon, I converse, and am merry with my friends; and when after three or four hours’ amusement, I would return to these speculations, they appear so cold, and strained, and ridiculous, that I cannot find in my heart to enter into them any farther. (*A Treatise of Human Nature*, Book I, Part IV, section VII)

The great subverter of Pyrrhonism or the excessive principles of skepticism is action, and employment, and the occupations of common life. These principles may flourish and triumph in the schools; where it is, indeed, difficult, if not impossible, to refute them. But as soon as they leave the shade, and by the presence of the real objects, which actuate our passions and sentiments, are put in opposition to the more powerful principles of our nature, they vanish like smoke, and leave the most determined skeptic in the same condition as other mortals . . . And though a Pyrrhonian may throw himself or others into a momentary amazement and confusion by his profound reasonings; the first and most trivial event in life will put to flight all his doubts and scruples, and leave him the same, in every point of action and speculation, with the philosophers of every other sect, or with those who never concerned themselves in any philosophical researches. When he awakes from his dream, he will be the first to join in the laugh against himself, and to confess that all his objections are mere amusement. (*An Enquiry Concerning Human Understanding*, Section XII, Part II)

The theory of knowledge we have presented explains why skeptics of various sorts have had such difficulties in sticking to their far-reaching skeptical conclusions “outside the study,” or even inside it when they are not thinking specifically about skeptical arguments and possibilities SK.

The skeptic’s arguments do show (but show only) that we don’t know the skeptic’s possibilities SK do not hold; and he is right that we don’t track the fact that SK does not hold. (If it were to hold, we would still think it didn’t.) However, the skeptic’s arguments don’t show we do not know other facts (including facts that entail not-SK) for we do track these other facts (and knowledge is not closed under known logical entailment.) Since we do track these other facts – you, for example, the fact that you are reading a book; I, the fact that I am writing on a page – and the skeptic tracks such facts too, it is not surprising that when he focuses on them, on his relationship to such facts, the skeptic finds it hard to remember or maintain his view that he does not know those facts. Only by shifting his attention back to his relationship to the (different) fact that not-SK, which relationship is not tracking, can he revive his skeptical belief and make it salient. However, this skeptical triumph is

evanescent, it vanishes when his attention turns to other facts. Only by fixating on the skeptical possibilities SK can he maintain his skeptical virtue; otherwise, unsurprisingly, he is forced to confess to sins of credulity.

Skepticism and the conditions for knowledge

We have considered how the skeptic’s argument from the skeptical possibilities SK plays off condition 3: if p weren’t true S wouldn’t believe that p . His argument gains its power by utilizing this condition (“but even if SK held, you still would believe it didn’t, so you do not know it doesn’t”); the deep intuitive force of the argument indicates that condition 3 (or something very much like it) is a necessary condition for knowledge. Similarly, are there any skeptical arguments or moves that play off condition 4: if p were true then S would believe that p (and wouldn’t believe that not- p)? If condition 3 specifies how belief somehow should vary with the truth of what is believed, condition 4 specifies how belief shouldn’t vary when the truth of what is believed does not vary. Condition 3 is a variation condition, condition 4 is an adherence condition. Both conditions together capture the notion that S (who actually truly believes p) would have a true belief that p . He wouldn’t have a false belief that p if p weren’t true (condition 3), and he would have a true belief that p if p were true (condition 4). Just as the skeptic argued earlier that the belief wouldn’t vary when it should, he also can argue that it would vary when it shouldn’t, concluding both times that we don’t have knowledge.

We would expect skeptical arguments playing off condition 4 to be less powerful and compelling than ones playing off 3. Condition 3 requires that we wouldn’t falsely believe p , and we can be led to worry not only whether we might but whether we do. While condition 4 requires that we would truly believe p (and wouldn’t falsely believe not- p), and though we might worry whether we might violate this, we need have no fear that we are – for we know we are believing p and are not believing not- p . Skeptical arguments playing off condition 4, unlike those with 3, cannot make us wonder also whether we violate the condition’s indicative version.

Condition 4 is an adherence condition, so the relevant doubts concern how securely you are tied to the truth. For many (most?) of the things p you

believe, if a group of people came and deceitfully told you not- p , you would believe them and stop believing p . (Relevant experiments frequently have been done by social psychologists.) So do you really know p ? If physicists told you that Newton's theory turns out to have been correct after all, wouldn't (or mightn't) you believe them? So do you really know Newtonian theory is false?

But, as before, the mere possibility of its being true while you do not believe it is not sufficient to show you don't actually know it. That possibility must be one that might arise. Call this possibility of p 's being true while you don't believe it: sk. (Lowercase "sk" is p 's being true and your not believing it, while capital SK is p 's being false and your believing p .) Possibility sk need not concern us when: if p were true, sk wouldn't hold; $p \rightarrow$ not-sk; sk is false throughout the first part of the p neighborhood of the actual world. It is fortunate for my knowing that p that there wouldn't be people who trick me, just as it is fortunate for my knowing I am in Emerson Hall that whatever would occur if I weren't there does not include people tricking or hypnotizing me into believing I am there.

Suppose I present a certain argument to someone who believes (truly) that p , and he is convinced by it and comes to believe not- p . Look how easily he can be moved from believing p to believing not- p . Suppose it happens that I do not present the argument to him, so he does not start to believe not- p , and he continues to believe p . Does he know that p ? Is it merely the case that his knowledge is insecure, or does such instability show it is not knowledge after all?

A skeptic might argue that for almost each p we (think we) know, there is an argument or happening that would get us to believe not- p even though p was true. We reply to this skeptic as before – the fact that some possible argument or happening would get us to believe not- p when p doesn't show that it is false that 4: if p were true then S would believe p and S wouldn't believe not- p . To show the falsity of 4, the skeptic would have to refer to something that might occur if p were true; if it wouldn't hold if p were true, what he refers to is irrelevant.

Among the arguments that get people to stop believing things are the skeptic's arguments themselves. These arguments often puzzle people, sometimes they get people to stop believing they know that p . They do not know that they know. Should we describe this as a case of people who

first know that they know but who, after hearing the skeptic's arguments, no longer know that they know because they no longer believe that they know (and knowledge entails belief)? Our present view is that such people did not know that they knew that p , even before hearing the skeptic. For their previous belief that they knew that p would vary when it shouldn't, so it violates condition 4. Similarly, some people who never have heard the skeptic's arguments would (if they heard them) become convinced that they don't know that p . It is pleasant to grant the skeptic a partial victory after all, one gained by the plausibility of his arguments, not their cogency. Because of the skeptical arguments, some people would falsely believe they don't know that p , and these people do not know they know it. The existence of skeptical arguments makes one type of skeptical conclusion (that we don't know we know things) true of some people – those the shoe fits have been wearing it.

Meno claimed he could speak eloquently about virtue until Socrates, torpedolike, began to question him. He did not know what virtue was, for Socrates' questions uncovered Meno's previously existing confusions. Even if it had been a sophist's questions that bewildered Meno, getting him to believe the opposite, what he previously had would not have been knowledge. Knowledge should be made of sterner stuff.³⁵

Thus, some skeptical arguments play off condition 3, others off condition 4. In addition to these conditions, our (full) account of knowledge formulates a condition about outweighing to cover the situation when multiple methods, not all satisfying 3 and 4, give rise to the belief. Do any skeptical arguments play off this outweighing condition? Here, presumably, would fit various attempts at unmasking the dominant sources of our belief as methods that do not track: faith, prejudice, self-interest, class-interest, deep psychological motives. The outweighing view involves subjunctives, but does anything here correspond to the skeptic's focusing upon a possibility that is so far out that it wouldn't occur, even if p were false? Perhaps the following is comparable. Recall that it was not necessary for the tracking method to win out against the combined opposed weight of all other methods; the person's belief merely had to vary with the verdict of the tracking method when the recommendations of every other way used to arrive at belief were held fixed. (It was only Case III in the chart that needed to be examined.) Any actual split in the verdict of nontracking methods will be

welcome support. The skeptic should not load the other methods against what tracking recommends, any more than they actually are; to suppose more counts as too far out.

Some skeptical arguments play off condition 3, some off condition 4, some (perhaps) off the out-weighing condition when multiple methods are involved. Still other skeptical arguments play off the methods themselves, off the fact that knowledge is gained via methods or ways of believing. In the situations when we are aware of what methods we are using, do we know we are using those methods? To decide whether we know this, according to condition 3 we must consider what we would believe if we weren't using the methods. Would we then still believe we were? If so, condition 3 is violated, and so we did not actually know we were using the methods.

Along this pathway lies trouble. For if we weren't using that method, the very method we use to track various facts – a situation we have to contemplate in applying condition 3 – who knows what we would believe about what methods we are using? That method *M* we are using to track various facts may be the very method via which we believe that we are using method *M*. This is likely if (and only if) *M* is described widely and deeply enough, for example, as the sum total of our (rational or effective) methods. But then, how are we to treat the question of what we would believe if we weren't using that method *M*, a question condition 3 pushes at us in order to decide if we know we are using *M*? “If I weren't using *M*, would I still believe I was?” What methods of believing am I left by this question? After all, condition 3 when fully formulated says: not-*p* and *S*, via *M*, comes to a belief about the truth of $p \rightarrow \text{not}-(S \text{ believes that } p)$. And the method *M* of condition 3 is the very one said to be actually utilized, in condition 2: *S* believes, via *M*, that *p*.

Yet now we face the situation where *S* believes of himself that he is applying method *M*, via an application of method *M* itself;³⁶ moreover, in this situation the statement *p*, which we are trying to decide whether *S* knows, is: *S* is using method *M*. The result of substituting this *p* in the full condition 3 is: If *S* weren't using method *M*, and *S*, via using *M*, were to decide about the truth of “*S* is using method *M*” then *S* would not believe “*S* is using method *M*.” But the antecedent of this subjunctive is supposing both that *S* is not using method *M* (this supposition is the not-*p* of the antecedent of condition 3) and that *S* is using method *M* (he uses this method in

3 to decide whether or not *p*, since that is the method via which, in condition 2, he actually believes *p*). We have no coherent way to understand this.³⁷

Yet if we cannot simply include the use of method *M* in determining what *S* would believe if he were not using *M*, neither can we simply suppose (for the purposes of condition 3) that *S* is using some other method to arrive at a belief about this matter. We saw earlier, in considering a range of examples, the great importance of holding the method fixed in deciding questions about knowledge. Recall the grandmother who sees her grandson visit her and so believes he is healthy and ambulatory; yet if he weren't ambulatory, other relatives would tell her he was fine to spare her anxiety and upset. She sees her grandson walking; does she know he is ambulatory? According to condition 3 we must ask what she would believe if he weren't ambulatory. If the method via which she believes is not held fixed, the answer will be wrong. True, if he weren't ambulatory, she would then believe he was (via hearing about him from other relatives). But the relevant question is: what would she believe if he weren't ambulatory and (as before) she saw him and spoke to him. Thus, to reach the correct answer about her knowledge, the method must be held fixed – that is one of the reasons why we introduced explicit reference to the method or way of believing.

How then are we to treat the question of whether the person knows he is using method *M*, when he believes he is via that very method *M*? If he knows he is, then his belief that he is tracks the fact that he is, and varies with that fact. To determine whether it so varies, we must look to the question of what he would believe if *p* were false, that is, if he weren't using method *M*. How are we to understand this question? It seems we must hold fixed the method *M* via which he believes, in order to reach the correct answer about knowledge (as is shown by the case of the grandmother), and that we cannot hold the method *M* fixed, for then we have the (apparently) incoherent supposition that he is applying the method to the situation where he is not using it, in order to determine whether or not he is – and this supposes that he both is and isn't using the method.

This problem does not arise when we know via another method that we are using some particular method; it arises only for our knowledge of our use of our deepest methods, though not for shallower specifications of these methods in specific

instances. Still, what should we say about our knowledge of these deepest methods or of the conditions in which we apply them. Do you know you are rational, do you know you are sane? If you were irrational or insane, mightn't you think you were rational and sane? Yes, but not by applying methods under (fixed) conditions of rationality and sanity. We cannot conclude simply that condition 3 is not satisfied so you don't know you are rational or sane; for that condition is not satisfied only when the method is allowed to vary. It would be best to be able coherently to discover whether or not that method is being used. I can use *M* to discover whether you are using *M* (if you weren't, I wouldn't believe, via *M*, that you were), or whether I was using *M* in the past (if I hadn't been, I wouldn't now believe, via *M*, that I had been). The difficulty is to make sense of saying that *M*, if currently used, would detect that it was not being used (if it weren't). And while I do not think this simply is incoherent, neither is it pellucidly clear.³⁸

Questions about knowing one is rational or sane need not depend on varying the method used. If what we have to go on as we apply methods is the appearance of rationality and sanity, then mightn't we appear sane and rational to ourselves even if we are not? So how do we know we are? We do have more to go on than how we appear to ourselves; there also is the agreement with others. Let us

leave aside the possibility that all those others also might be insane and irrational, or be engaged in a plot to convince me (falsely) that I was rational and sane. Neither of these is what (actually) would or might occur if I weren't rational or sane. Might an insane and irrational person also be mistaken about whether others are agreeing with him, though, interpreting their disagreement as concord? If a person were insane or irrational in this way then others would appear (to him) to agree with him, and so he would appear sane and rational to himself. Things would appear qualitatively indistinguishable to him from the situation where he rationally and sanely judges the world. There appears to be no shift in method here, at least insofar as how using the method is experienced internally by the user. Do you know, then, that you are not in that particular skeptical situation SK? Perhaps not, but (as before) from our not knowing that particular not-SK it does not follow that we don't know other things, including that we are being sane and rational in particular situations in particular ways. For if we weren't, we wouldn't believe we were; if we weren't then sane and rational in those particular ways, what would or might obtain is not this skeptic's possibility SK. These points emerge even more clearly if we consider positions skeptical not about (almost) all knowledge in general, but about particular kinds of knowledge.

Notes

1 Despite some demurrals in the literature, there is general agreement that conditions 1 and 2 are necessary for knowledge. (For some recent discussions, see D. M. Armstrong, *Belief, Truth and Knowledge* (Cambridge: Cambridge University Press, 1973), ch. 10; Keith Lehrer, *Knowledge* (Oxford: Oxford University Press, 1974), chs. 2, 3.) I shall take for granted that this is so, without wishing to place very much weight on its being belief that is the precise cognitive attitude (as opposed to thinking it so, accepting the statement, and so on) or on the need to introduce truth as opposed to formulating the first condition simply as: *p*.

I should note that our procedure here does not stem from thinking that every illuminating discussion of an important philosophical notion must present (individually) necessary and (jointly) sufficient conditions.

2 Below, we discuss further the case where though the fact that *p* causes the person's belief that *p*, he would

believe it anyway, even if it were not true. I should note here that I assume bivalence throughout this chapter, and consider only statements that are true if and only if their negations are false.

3 See Robert Stalnaker, "A Theory of Conditionals," in N. Rescher (ed.), *Studies in Logical Theory* (Oxford: Basil Blackwell, 1968); David Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973); and Jonathan Bennett's critical review of Lewis, "Counterfactuals and Possible Worlds," *Canadian Journal of Philosophy* IV, 2 (Dec. 1974), pp. 381-402.

Our purposes require, for the most part, no more than an intuitive understanding of subjunctives. However, it is most convenient to examine here some further issues, which will be used once or twice later. Lewis's account has the consequence that $p \rightarrow q$ whenever *p* and *q* are both true; for the possible world where *p* is true that is closest to the actual world is the actual world itself, and in that

world q is true. We might try to remedy this by saying that when p is true, $p \rightarrow q$ is true if and only if q is true in all p worlds closer (by the metric) to the actual world than is any not- p world. When p is false, the usual accounts hold that $p \rightarrow q$ is true when q holds merely in the closest p worlds to the actual world. This is too weak, but how far out must one go among the p worlds? A suggestion parallel to the previous one is: out until one reaches another not- p world (still further out). So if q holds in the closest p world w_1 but not in the p world w_2 , even though no not- p world lies between w_1 and w_2 , then (under the suggestion we are considering) the subjunctive is false. A unified account can be offered for subjunctives, whatever the truth value of their antecedents. The p neighborhood of the actual world A is the closest p band to it; that is, w is in the p neighborhood of the actual world if and only if p is true in w and there are no worlds $w^{\bar{p}}$ and w^p such that not- p is true in $w^{\bar{p}}$ and p is true in w^p , and $w^{\bar{p}}$ is closer to A than w is to A , and w^p is at least as close to A as $w^{\bar{p}}$ is to A . A subjunctive $p \rightarrow q$ is true if and only if q is true throughout the p neighborhood of the actual world.

If it is truly a random matter which slit a photon goes through, then its going through (say) the right slit does not establish the subjunctive: if a photon were fired at that time from that source it would go through the right-hand slit. For when p equals A a photon is fired at that time from that source, and q equals the photon goes through the right-hand slit, q is not true everywhere in the p neighborhood of the actual world.

This view of subjunctives within a possible-worlds framework is inadequate if there is no discrete p band of the actual world, as when for each positive distance from the actual world A , there are both p worlds and not- p worlds so distant. Even if this last is not generally so, many p worlds that interest us may have their distances from A matched by not- p worlds. Therefore, let us redefine the relevant p band as the closest spread of p worlds such that there is no not- p world intermediate in distance from A to two p worlds in the spread unless there is also another p world in the spread the very same distance from A . By definition, it is only p worlds in the p band, but some not- p worlds may be equidistant from A .

Though this emendation allows us to speak of the closest spread of p worlds, it no longer is so clear which worlds in this p band subjunctives (are to) encompass. We have said it is not sufficient for the truth of $p \rightarrow q$ that q hold in that one world in the p band closest to the actual world. Is it necessary, as our first suggestion has it, that q hold in all the p worlds in the closest p band to the actual world? Going up until the first "pure" stretch of not- p worlds is no longer as natural a line to draw as when we imagined "pure" p neighborhoods. Since there already are some not- p worlds the same distance from A as some members of

the p band, what is the special significance of the first unsullied not- p stretch? There seems to be no natural line, though, coming before this stretch yet past the first p world. Perhaps nothing stronger can be said than this: $p \rightarrow q$ when q holds for some distance out in the closest p band to the actual world; that is, when all the worlds in this first part of that closest p band are q . The distance need not be fixed as the same for all subjunctives, although various general formulas might be imagined, for example, that the distance is a fixed percentage of the width of the p band.

I put forth this semantics for subjunctives in a possible-worlds framework with some diffidence, having little inclination to pursue the details. Let me emphasize, though, that this semantics does not presuppose any realist view that all possible worlds obtain. I would hope that into this chapter's subjunctively formulated theoretical structure can be plugged (without too many modifications) whatever theory of subjunctives turns out to be adequate, so that the theory of knowledge we formulate is not sensitive to variations in the analysis of subjunctives. In addition to Lewis and Stalnaker cited above, see Ernest W. Adams, *The Logic of Conditionals* (Dordrecht: Reidel, 1975); John Pollock, *Subjunctive Reasoning* (Dordrecht: Reidel, 1976); J. H. Sobel, "Probability, Chance and Choice" (unpublished book manuscript); and a forthcoming book by Yigal Kvat.

- 4 If the possible-worlds formalism is used to represent counterfactuals and subjunctives, the relevant worlds are not those p worlds that are closest or most similar to the actual world, unless the measure of closeness or similarity is: what would obtain if p were true. Clearly, this cannot be used to explain when subjunctives hold true, but it can be used to represent them. Compare utility theory which represents preferences but does not explain them. Still, it is not a trivial fact that preferences are so structured that they can be represented by a real-valued function, unique up to a positive linear transformation, even though the representation (by itself) does not explain these preferences. Similarly, it would be of interest to know what properties hold of distance metrics which serve to represent subjunctives, and to know how subjunctives must be structured and interrelated so that they can be given a possible worlds representation. (With the same one space serving for all subjunctives?)

One further word on this point. Imagine a library where a cataloguer assigns call numbers based on facts of sort F . Someone, perhaps the cataloguer, then places each book on the shelf by looking at its call number, and inserting it between the two books whose call numbers are most nearly adjacent to its own. The call number is derivative from facts of type F , yet it plays some explanatory role, not merely a representational one. "Why is this book located

precisely there? Because of its number." Imagine next another library where the person who places books on the shelves directly considers facts of type F, using them to order the books and to interweave new ones. Someone else might notice that this ordering can be represented by an assignment of numbers, numbers from which other information can be derived as well, for example, the first letter of the last name of the principal author. But such an assigned number is no explanation of why a book in this library is located between two others (or why its author's last name begins with a certain letter). I have assumed that utility numbers stand to preferences, and closeness or similarity measures stand to subjunctives, as the call numbers do to the books, and to the facts of type F they exhibit, in the second library.

- 5 G. C. Stine, "Skepticism, Relevant Alternatives and Deductive Closure," *Philosophical Studies* 29 (1976), p. 252, who attributes the example to Carl Ginet.
- 6 This last remark is a bit too brisk, for that account might use a subjunctive criterion for when an alternative q to p is relevant (namely, when if p were not to hold, q would or might), and utilize some further notion of what it is to rule out relevant alternatives (for example, have evidence against them), so that it did not turn out to be equivalent to the account we offer.
- 7 More accurately, since the truth of antecedent and consequent is not necessary for the truth of the subjunctive either, 4 says something different from 1 and 2.
- 8 I experimented with some other conditions which adequately handled this as well as some other problem cases, but they succumbed to further difficulties. Though much can be learned from applying those conditions, presenting all the details would engage only the most masochistic readers. So I simply will list them, each at one time a candidate to stand alone in place of condition 4.
 - (a) S believes that $\text{not-}p \rightarrow \text{not-}p$.
 - (b) S believes that $\text{not-}p \rightarrow \text{not-}p$ or it is through some other method that S believes $\text{not-}p$. (Methods are discussed in the next section.)
 - (c) (S believes p or S believes $\text{not-}p$) \rightarrow $\text{not-}(S$ believes p , and $\text{not-}p$ holds) and $\text{not-}(S$ believes $\text{not-}p$, and p holds).
 - (d) $\text{not-}(S$ believes that p) \rightarrow $\text{not-}(p$ and S believes that $\text{not-}p$).
 - (e) $\text{not-}(p$ and S believes that p) \rightarrow $\text{not-}(\text{not-}p$ and S believes that p or p and S believes that $\text{not-}p$).
- 9 Gilbert Harman, *Thought* (Princeton: Princeton, University Press, 1973), ch. 9, pp. 142–54.
- 10 What if the situation or world where he too hears the later false denials is not so close, so easily occurring? Should we say that everything that prevents his

hearing the denial easily could have not happened, and does not in some close world?

- 11 This reformulation introduces an apparent asymmetry between the consequents of conditions 3 and 4. Since we have rewritten 4 as

$p \rightarrow$ S believes that p and $\text{not-}(S$ believes that $\text{not-}p$),
why is 3 not similarly rewritten as

$\text{not-}p \rightarrow \text{not-}(S$ believes that p) and S believes that $\text{not-}p$?

It is knowledge that p we are analyzing, rather than knowledge that $\text{not-}p$. Knowledge that p involves a stronger relation to p than to $\text{not-}p$. Thus, we did not first write the third condition for knowledge of p as: $\text{not-}p \rightarrow$ S believes that $\text{not-}p$; also the following is not true: S knows that $p \rightarrow (\text{not-}p \rightarrow$ S knows that $\text{not-}p$).

Imagine that someone S knows whether or not p , but it is not yet clear to us which he knows, whether he knows that p or knows that $\text{not-}p$. Still, merely given that S knows that—, we can say:

$\text{not-}p \rightarrow \text{not-}(S$ believes that p)
 $p \rightarrow \text{not-}(S$ believes that $\text{not-}p$).

Now when the blank is filled in, either with p or with $\text{not-}p$, we have to add S's believing it to the consequent of the subjunctive that begins with it. That indicates which one he knows. Thus, when it is p that he knows, we have to add to the consequent of the second subjunctive (the subjunctive that begins with p): S believes that p . We thereby transform the second subjunctive into:

$p \rightarrow \text{not-}(S$ believes that $\text{not-}p$) and S believes that p .

Except for a rearrangement of which is written first in the consequent, this is condition 4. Knowledge that p especially tracks p , and this special focus on p (rather than $\text{not-}p$) gets expressed in the subjunctive, not merely in the second condition.

There is another apparent asymmetry in the antecedents of the two subjunctives 3 and 4, not due to the reformulation. When actually p is true and S believes that p , condition 4 looks some distance out in the p neighborhood of the actual world, while condition 3 looks some distance out in the $\text{not-}p$ neighborhood, which itself is farther away from the actual world than the p neighborhood. Why not have both conditions look equally far, revising condition 3 to require merely that the closest world in which p is false yet S believes that p be some distance from the actual world. It then would parallel condition 4, which says that the closest world in which p yet p is not believed is some distance away from the actual world. Why should condition 3 look farther from the actual world than condition 4 does?

However, despite appearances, both conditions look at distance symmetrically. The asymmetry is

caused by the fact that the actual world, being a p world, is not symmetrical between p and not- p . Condition 3 says that in the closest not- p world, not-(S believes that p), and that this “not-(S believes that p)” goes out through the first part of the not- p neighborhood of the actual world. Condition 4 says that in the closest p world, S believes that p , and that this “S believes that p ” goes out through the first part of the p neighborhood of the actual world. Thus the two conditions are symmetrical; the different distances to which they extend stems not from an asymmetry in the conditions but from one in the actual world – it being (asymmetrically) p .

- 12 D. M. Armstrong, *Belief, Truth and Knowledge* (Cambridge: Cambridge University Press, 1973), p. 209; he attributes the case to Gregory O’Hair.
- 13 Some may hold the father is made more sure in his belief by courtroom proof; and hold that the father knows because his degree of assurance (though not his belief) varies subjunctively with the truth.
- 14 If there is no other such method M_1 via which S believes that p , the second clause is vacuously true.

Should we say that no other method used outweighs M , or that M outweighs all others? Delicate questions arise about situations where the methods tie, so that no subjunctive holds about one always winning over the other. It might seem that we should require that M outweigh (and not merely tie) the other methods; but certain ways of resolving the ties, such as not randomly deciding but keeping judgment suspended, might admit knowledge when a true belief is arrived at via a tracking method M which is not outweighed yet also doesn’t (always) outweigh the others present. There is no special need to pursue the details here; the outweighing condition should be read here and below as a vague one, residing somewhere in the (closed) interval between “outweighs” and “not outweighed”, but not yet precisely located. This vagueness stands independently of the refinements pursued in the text immediately below.

- 15 When a belief is overdetermined or jointly produced by three methods, where only the first satisfies conditions 3 and 4, the question becomes: what does the person believe when M_1 recommends believing not- p while the two others each recommend believing p ? Notice also that in speaking of what would happen in Case III we are imposing a subjunctive condition; if there is no “would” about it, if in each instance of a Case III situation it is determined at random which method outweighs which, then that will not be sufficient for knowledge, even though sometimes M_1 wins out.

It is worrisome that in weakening our initial description of outweighing by looking to Case III but not to Case II, we seem to give more weight to condition 3 for tracking than to condition 4. So we should be ready to reconsider this weakening.

- 16 For example, in the case of the father who believes on faith that his son is innocent and sees the courtroom demonstration of innocence, does the father use two methods, faith and courtroom demonstration, the second of which does satisfy conditions 3–4 while the first (which outweighs it) does not satisfy 3–4; or does the father use only one method which doesn’t satisfy 3–4, namely: believe about one’s son whatever the method of faith tells one, and only if it yields no answer, believe the result of courtroom demonstration? With either mode of individuation, knowledge requires the negative existentially quantified statement (that there is no method . . .) somewhere, whether in specifying the method itself or in specifying that it is not outweighed.
- 17 One suspects there will be some gimmick whereby whenever p is truly believed a trivial method M can be specified which satisfies conditions 3 and 4. If so, then further conditions will have to be imposed upon M , in addition to the dispositional condition. Compare the difficulties encountered in the literature on specifying the relevant reference class in probabilistic inference and explanation; see Henry Kyburg, *Probability and the Logic of Rational Belief* (Middletown: Wesleyan University Press, 1961), ch. 9; C. G. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 394–405; also his “Maximal Specificity and Lawlikeness in Probabilistic Explanation,” *Philosophy of Science* 35 (1968), pp. 116–33.
- 18 See Ludwig Wittgenstein, *On Certainty* (Oxford: Basil Blackwell, 1969), 83, 94, 102–10, 140–4, 151–2, 162–3, 166, 411, 419, 472–5.
- 19 There is an immense amount of literature concerning skepticism. See, for example, Sextus Empiricus, *Writings* (4 vols, Loeb Classical Library, Cambridge: Harvard University Press); Richard Popkin, *History of Skepticism from Erasmus to Descartes* (rev. edn, New York: Humanities Press, 1964); Arne Naess, *Skepticism* (New York: Humanities Press, 1968); René Descartes, *Meditations on First Philosophy* (New York: Liberal Arts Press, 1960); G. E. Moore, “Proof of an External World,” “Certainty,” and “Four Forms of Scepticism,” this vol., chs 2, 3 and 4, and “A Defense of Common Sense,” in his *Philosophical Papers* (Allen and Unwin, London, 1959); J. L. Austin, “Other Minds” in his *Philosophical Papers* (Oxford University Press, 1961); Wittgenstein, *On Certainty*; Keith Lehrer, “Why Not Skepticism?” (in Swain and Pappas (eds), *Essays on Knowledge and Justification*, pp. 346–63); Peter Unger, *Ignorance* (Oxford University Press, 1975), pp. 7–24; Michael Slote, *Reason and Skepticism* (London: Allen and Unwin, 1970); Roderick Firth, “The Anatomy of Certainty,” *Philosophical Review* 76 (1967), pp. 3–27; Thompson Clarke, “The Legacy of Skepticism,” *Journal of Philosophy* 69 (1972), pp. 754–69; Stanley Cavell,

The Claim of Reason (Oxford University Press, 1979).

- 20 From the perspective of explanation rather than proof, the extensive philosophical discussion, deriving from Charles S. Peirce, of whether the skeptic's doubts are real is beside the point. The problem of explaining how knowledge is possible would remain the same, even if no one ever claimed to doubt that there was knowledge.
- 21 Subjunctives with actually false antecedents and actually true consequents have been termed by Goodman *semi-factuals*. R is the semi-factual: not- $p \rightarrow S$ believes p .
- 22 Should one weaken condition 3, so that the account of knowledge merely denies the opposed subjunctive R ? That would give us: not-(not- $p \rightarrow S$ believes p). This holds when 3 does not, in situations where if p were false, S might believe p , and also might not believe it. The extra strength of 3 is needed to exclude these as situations of knowledge.
- 23 Though it does show the falsity of the corresponding entailment, "not- p entails not-(S believes that p)."
- 24 If a person is to know that SK doesn't hold, then condition 3 for knowledge must be satisfied (with "SK doesn't hold" substituted for p). Thus, we get
- (3) not-(SK doesn't hold) \rightarrow not-(S believes that SK doesn't hold).

Simplifying the antecedent, we have

- (3) SK holds \rightarrow not-(S believes that SK doesn't hold).

The skeptic has chosen a situation SK such that the following is true of it:

SK holds $\rightarrow S$ believes that SK doesn't hold.

Having the same antecedent as 3 and a contradictory consequent, this is incompatible with 3. Thus, condition 3 is not satisfied by the person's belief that SK does not hold.

- 25 Descartes presumably would refute the tank hypothesis as he did the demon hypothesis, through a proof of the existence of a good God who would not allow anyone, demon or psychologist, permanently to deceive us. The philosophical literature has concentrated on the question of whether Descartes can prove this (without begging the question against the demon hypothesis). The literature has not discussed whether even a successful proof of the existence of a good God can help Descartes to conclude he is not almost always mistaken. Might not a good God have his own reasons for deceiving us; might he not deceive us temporarily – a period which includes all of our life thus far (but not an afterlife)? To the question of why God did not create us so that we never would make any errors, Descartes answers that the motives of God are inscrutable to us. Do

we know that such an inscrutable God could not be motivated to allow another powerful "demon" to deceive and dominate us?

Alternatively, could not such a good God be motivated to deceive itself temporarily, even if not another? (Compare the various Indian doctrines designed to explain our ignorance of our own true nature, that is, Atman-Brahman's or, on another theory, the purusha's nature.) Whether from playfulness or whatever motive, such a good God would temporarily deceive itself, perhaps even into thinking it is a human being living in a material realm. Can we know, via Descartes' argument, that this is not our situation? And so forth.

These possibilities, and others similar, are so obvious that some other explanation, I mean the single-minded desire to refute skepticism, must be given for why they are not noticed and discussed.

Similarly, one could rescrutinize the *cogito* argument. Can "I think" only be produced by something that exists? Suppose Shakespeare had written for Hamlet the line, "I think, therefore I am," or a fiction is written in which a character named Descartes says this, or suppose a character in a dream of mine says this; does it follow that they exist? Can someone use the *cogito* argument to prove he himself is not a fictional or dream character? Descartes asked how he could know he wasn't dreaming; he also should have asked how he could know he wasn't dreamed. See further my fable "Fiction," *Ploughshares* 6, 3 (Oct. 1980).

- 26 I say almost everything, because there still could be some true beliefs such as "I exist." More limited skeptical possibilities present worlds doxically identical to the actual world in which almost every belief of a certain sort is false, for example, about the past, or about other people's mental states.
- 27 Let $w_1 \dots, w_n$ be worlds doxically identical to the actual world for S . He doesn't know he is not in w_1 , he doesn't know he is not in $w_2 \dots$; does it follow that he doesn't know he is in the actual world w_A or in one very much like it (in its truths)? Not if the situation he would be in if the actual world w_A did not obtain wasn't one of the doxically identical worlds; if the world that then would obtain would show its difference from the actual one w_A , he then would not believe he was in w_A .

However, probably there are some worlds not very different from the actual world (in that they have mostly the same truths) and even doxically identical to it, which might obtain if w_A did not. In that case, S would not know he was in w_A specified in all its glory. But if we take the disjunction of these harmless worlds (insofar as drastic skeptical conclusions go) doxically identical with w_A then S will know that the disjunction holds. For if it didn't, he would notice that.

- 28 This argument proceeds from the fact that floating in the tank is incompatible with being at X. Another form of the skeptic's argument, one we shall consider later, proceeds from the fact that floating in the tank is incompatible with knowing you are at X (or almost anything else).
- 29 Note that I am not denying that $Kp \ \& \ K(p \prec q) \rightarrow$ Believes q .
- 30 Here again I assume a limited readership, and ignore possibilities such as those described in James Blish, *Cities in Flight*.
- 31 Thus, the following is not a deductively valid form of inference.

$p \prec q$ (and S knows this)
 $\text{not-}p \rightarrow \text{not-}(S \text{ believes that } p)$
 Therefore, $\text{not-}q \rightarrow \text{not-}(S \text{ believes that } q)$.

Furthermore, the example in the text shows that even the following is not a deductively valid form of inference.

$p \prec q$ (and S knows this)
 $\text{not-}p \rightarrow \text{not-}(S \text{ believes that } p)$
 Therefore, $\text{not-}q \rightarrow \text{not-}(S \text{ believes that } p)$.

Nor is this one deductively valid:

$p \prec q$
 $\text{not-}q \rightarrow r$
 Therefore, $\text{not-}p \rightarrow r$.

- 32 Does this same consequence of nonclosure under known logical implication follow as well from condition 4: $p \rightarrow S$ believes that p ? When p is not actually true, condition 4 can hold of p yet not of a q known to be entailed by p . For example, let p be the (false) statement that I am in Antarctica, and let q be the disjunction of p with some other appropriate statement; for example, let q be the statement that I am in Antarctica or I lost some object yesterday though I have not yet realized it. If p were true I would know it, p entails q , yet if q were true I wouldn't know it, for the way it would be true would be by my losing some object without yet realizing it, and if that happened I would not know it.

This example to show that condition 4 is not closed under known logical implication depends on the (actual) falsity of p . I do not think there is any suitable example to show this in the case where p is true, leaving aside the trivial situation when the person simply does not infer the entailed statement q .

- 33 Suppose some component of the condition, call it C, also speaks of some cases when p is false, and when q is false; might it then provide "varies with," even though C is preserved under known logical implication, and is transmitted from p to q when p entails q and is known to entail q ? If this condition C speaks of some cases where $\text{not-}p$ and of some cases where $\text{not-}q$, then C will be preserved under known logical implication if, when those cases of $\text{not-}p$ satisfy it,

and p entails q , then also those cases of $\text{not-}q$ satisfy it. Thus, C seems to speak of something as preserved from some cases of $\text{not-}p$ to some cases of $\text{not-}q$, which is preservation in the reverse direction to the entailment involving these, from $\text{not-}q$ to $\text{not-}p$. Thus, a condition that is preserved under known logical implication and that also provides some measure of "varies with" must contain a component condition saying that something interesting (other than falsity) is preserved in the direction opposite to the logical implication (for some cases); and moreover, that component itself must be preserved in the direction of the logical implication because the condition including it is. It would be interesting to see such a condition set out.

- 34 Reading an earlier draft of this chapter, friends pointed out to me that Fred Dretske already had defended the view that knowledge (as one among many epistemic concepts) is not closed under known logical implication. (See his "Epistemic Operators," *Journal of Philosophy* 67 (1970), pp. 1007-23.) Furthermore, Dretske presented a subjunctive condition for knowledge (in his "Conclusive Reason," *Australasian Journal of Philosophy* 49, (1971), pp. 1-22), holding that S knows that p on the basis of reasons R only if: R would not be the case unless p were the case. Here Dretske ties the evidence subjunctively to the fact, and the belief based on the evidence subjunctively to the fact through the evidence. (Our account of knowledge has not yet introduced or discussed evidence or reasons at all. While this condition corresponds to our condition 3, he has nothing corresponding to 4.) So Dretske has hold of both pieces of our account, subjunctive and nonclosure, and he even connects them in a passing footnote (*Journal of Philosophy* 67, p. 1019, n. 4), noticing that any account of knowledge that relies on a subjunctive conditional will not be closed under known logical implication. Dretske also has the notion of a relevant alternative as "one that might have been realized in the existing circumstances if the actual state of affairs had not materialized" (p. 1021), and he briefly applies all this to the topic of skepticism (pp. 1015-16), holding that the skeptic is right about some things but not about others.

It grieves me somewhat to discover that Dretske also had all this, and was there first. It raises the question, also, of why these views have not yet had the proper impact. Dretske makes his points in the midst of much other material, some of it less insightful. The independent statement and delineation of the position here, without the background noise, I hope will make clear its many merits.

After Goldman's paper on a causal theory of knowledge (in *Journal of Philosophy* 64 (1967)), an idea then already "in the air," it required no great leap to consider subjunctive conditions. Some two months after the first version of this chapter was

- written, Goldman himself published a paper on knowledge utilizing counterfactuals ("Discrimination and Perceptual Knowledge," *Journal of Philosophy* 78 (1976), pp. 771–91), also talking of relevant possibilities (without using the counterfactuals to identify which possibilities are relevant); and Shope's survey article has called my attention to a paper of L. S. Carrier ("An Analysis of Empirical Knowledge," *Southern Journal of Philosophy* 9 (1971), pp. 3–11) that also used subjunctive conditions including our condition 3. Armstrong's reliability view of knowledge (*Belief, Truth and Knowledge*, pp. 166, 169) involved a lawlike connection between the belief that p and the state of affairs that makes it true. Clearly, the idea is one whose time has come.
- 35 Is it a consequence of our view that of two people who know p , each believing he knows p and satisfying condition 3 for knowing he knows p , one may know he knows and the other not, because (although identical in all other respects) the second might encounter skeptical arguments while the first some-
- how lives hermetically sealed from the merest brush with them?
- 36 Our task now is not to wonder whether it is legitimate to use M to reach a belief that M is being used. What, after all, is the alternative? Presumably, an infinite regress of methods, or a circle, or reaching a method which is used but either is not believed to be used, or is believed to be though not via any method or way of believing.
- 37 Similar questions arise about our knowledge of other statements such that if they were false, we would not be using the methods via which we know they are true, for example, "there are eyes," "I am alive," "I am sentient," perhaps "I sometimes am tracking something."
- 38 Should we say for these cases discussed in the text that condition 3 does not apply, so that, as in the previous case of necessary truths, the whole weight of tracking devolves upon condition 4? The issue then simply turns on whether in similar situations where the person uses method M , he also would believe he does.

PART III

Contemporary Foundationalism and Coherentism

Introduction

The Pyrrhonian problematic can be formulated as follows. One can be justified in believing that p only if one has a reason to believe that p . But if a proposition that q is one's reason to believe that p , it can provide justification only if it is a good reason – that is, only if it, too (or they, if there is more than one reason), is something one is justified in believing. This leaves three possibilities for any tree of justification: (1) all its branches terminate; (2) at least one of its branches contains a loop; (3) at least one of its branches is infinite. Thus, we have the three traditional theories of justification: foundationalism, coherentism, and the rather less popular infinitism. To be complete, there is a fourth option not mentioned, namely that skepticism is true and there are no trees of justification, for no one is ever justified in believing anything.

This description of the Pyrrhonian problematic corresponds closely to the way Roderick Chisholm sees the epistemological terrain. Faced with these options, he chooses foundationalism. This means admitting that there are some propositions that we are justified in believing but for which we lack reasons in the form of further propositions we are justified in believing. Chisholm fully embraces this consequence. Having made the statement "There lies a key," if one is asked "What is your justification for thinking *that*?" one must provide an answer, but eventually, in the chain of questions, a claim about one's present experience will be challenged: "What is your justification for thinking you have such-and-such experience?" To this, Chisholm thinks one can do no better than to answer "My justification is that I have such-and-such experience." Similarly, faced with a challenge to a claim regarding one's present

belief that p , one must repeat oneself, saying "My justification for thinking I believe that p is that I believe that p ."

Are, then, these foundational beliefs, these parts of "the given," themselves justified? Chisholm declines to give us an answer, saying that there are only two essential points: these beliefs, whether they are justified or not, justify other beliefs, and reports expressing these beliefs provide stopping places in the dialectic of asking for and giving reasons.

Wilfrid Sellars attacks the doctrine of the given precisely on the issue of the epistemological status of these foundational beliefs. If there is knowledge that is unsupported by further knowledge, as Chisholm would have to acknowledge, then reports of this knowledge, like reports of any piece of knowledge, must have authority. But a report can have authority only if the person making it in some sense recognizes its authority. Thus, even in the case of my knowledge that what I see before me is green, my report "This is green" must have authority that I recognize. Moreover, in this case, the authority can only lie in the reliable connection between the production of tokens of "This is green" and the presence of green objects. So if I am to know through observation that what I see is green, I must recognize the truth of this generalization. How, then, do I know the truth of the generalization? My present knowledge is based on memory knowledge of instances of it. What of my knowledge of these instances? Are we headed for a regress? No, answers Sellars, for although I have such memory knowledge, the experiential beliefs from which these memories are derived need not have been pieces of

Introduction

knowledge. (Presumably, these pieces of memory knowledge, too, have authority that is also recognized by the subject in the form of a belief that reports of "This was green and I experienced it to be so" co-vary with actual past encounters with green objects.)

Sellars's account rules out Chisholm's given. Formulated in the language of beliefs rather than of reports: if all knowledge of particular matters of fact, including even observation knowledge, depends on general knowledge, and if in turn this general knowledge itself depends on knowledge of particular matters of fact, then empirical knowledge has no foundation.

Ernest Sosa, in his contribution, clarifies the alternatives of coherentism and foundationalism, of the raft and the pyramid. Traditional foundationalism and coherentism alike are committed to a kind of "formal" foundationalism, which holds that epistemic conditions supervene on non-epistemic conditions in a way that can be specified in general, perhaps recursively. Formal foundationalism, according to Sosa, derives its plausibility from the claims that epistemic conditions are normative and that all normative conditions are supervenient. If a state of affairs is good, it must be good because it is a state of pleasure, or because it is a state of desire satisfaction, etc. It cannot be

barely good or good ultimately owing merely to the goodness of some other state(s). So, too, if a belief is justified, some non-epistemic condition must account for its justification. Sosa goes on to argue that the thesis of formal foundationalism conflicts with internalist theories of justification (perhaps such as Sellars's). If one's justification for believing that p is fixed ultimately by non-epistemic facts, then such justification cannot in every case also require the possession of further justified beliefs.

In the final selection of this section, Donald Davidson proposes a new coherentist theory, based on conclusions about meaning and content. Meaning, coherence, and truth, he argues, are internally connected. One's meaning what one does by one's words and in one's thoughts depends on one's being interpretable as a coherent (indeed a rational) believer most of whose beliefs are true. Since a fully informed or omniscient interpreter would also interpret any believer as having mostly true beliefs, it follows that all believers, ourselves included, have mostly true beliefs. This would establish a further connection to justification, according to Davidson, for in seeing that most of our beliefs are true, we gain a presumptive reason in favor of retaining any arbitrary one of them.

Further Reading

- Alston, William, "Two Types of Foundationalism," in *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989), pp. 19–38.
- Audi, Robert, *The Structure of Justification* (New York: Cambridge University Press, 1993).
- Bender, J. (ed.), *The Current State of the Coherence Theory* (Dordrecht: Kluwer Academic Publishers, 1989).
- BonJour, Laurence, *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Chisholm, Roderick, *The Foundations of Knowing* (Minneapolis: University of Minnesota Press, 1982).
- , *Theory of Knowledge* (Englewood Cliffs: Prentice-Hall, 1966, 2nd edn 1977, 3rd edn 1989).
- Haack, Susan, *Evidence and Inquiry: Towards Reconstruction in Epistemology* (Oxford: Blackwell, 1993).
- Lehrer, Keith, *Theory of Knowledge* (Boulder, CO: Westview Press, 1990).
- Lewis, C. I., *An Analysis of Knowledge and Valuation* (LaSalle, IL: Open Court Publishing Company, 1946).
- Plantinga, Alvin, *Warrant: The Current Debate* (Oxford: Oxford University Press, 1993).
- Rescher, Nicholas, *Methodological Pragmatism* (New York: New York University Press, 1977).
- Sellars, Wilfrid, "Empiricism and the Philosophy of Mind," reprinted in Sellars, *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963).
- , "Givenness and Explanatory Coherence," *Journal of Philosophy* 70 (1973), pp. 612–24.
- Sosa, Ernest, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991).

The Myth of the Given

Roderick M. Chisholm

I. The doctrine of "the given" involved two theses about our knowledge. We may introduce them by means of a traditional metaphor:

- (A) The knowledge which a person has at any time is a structure or edifice, many parts and stages of which help to support each other, but which as a whole is supported by its own foundation.

The second thesis is a specification of the first:

- (B) The foundation of one's knowledge consists (at least in part) of the apprehension of what have been called, variously, "sensations," "sense-impressions," "appearances," "sensa," "sense-qualia," and "phenomena"

These phenomenal entities, said to be at the base of the structure of knowledge, are what was called "the given." A third thesis is sometimes associated with the doctrine of the given, but the first two theses do not imply it. We may formulate it in the terms of the same metaphor:

- (C) The *only* apprehension which is thus basic to the structure of knowledge is our apprehension of "appearances" (etc.) – our apprehension of the given.

Theses (A) and (B) constitute the "doctrine of the given"; thesis (C), if a label were necessary, might

Originally published in R. Chisholm, *Philosophy* (Englewood Cliffs, NJ: Prentice-Hall, 1964), pp. 261–86

be called "the phenomenalist version" of the doctrine. The first two theses are essential to the empirical tradition in Western philosophy. The third is problematic for traditional empiricism and depends in part, but only in part, upon the way in which the metaphor of the edifice and its foundation is spelled out.

I believe it is accurate to say that, at the time at which our study begins, most American epistemologists accepted the first two theses and thus accepted the doctrine of the given. The expression "the given" became a term of contemporary philosophical vocabulary partly because of its use by C. I. Lewis in his *Mind and the World-Order* (Scribner, 1929). Many of the philosophers who accepted the doctrine avoided the expression because of its association with other more controversial parts of Lewis's book – a book which might be taken (though mistakenly, I think) also to endorse thesis (C), the "phenomenalist version" of the doctrine. The doctrine itself – theses (A) and (B) – became a matter of general controversy during the period of our survey.

Thesis (A) was criticized as being "absolute" and thesis (B) as being overly "subjective." Both criticisms may be found in some of the "instrumentalist" writings of John Dewey and philosophers associated with him. They may also be found in the writings of those philosophers of science ("logical empiricists") writing in the tradition of the Vienna Circle. (At an early stage of this tradition, however, some of these same philosophers seem to have accepted all three theses.) Discussion became entangled in verbal confusions – especially in connection with the uses of such terms as "doubt," "certainty," "appearance,"

and “immediate experience.” Philosophers, influenced by the work that Ludwig Wittgenstein had been doing in the 1930s, noted such confusions in detail, and some of them seem to have taken the existence of such confusions to indicate that (A) and (B) are false.¹ Many have rejected both theses as being inconsistent with a certain theory of thought and reference; among them, in addition to some of the critics just referred to, we find philosophers in the tradition of nineteenth-century “idealism.”

Philosophers of widely diverging schools now believe that “the myth of the given” has finally been dispelled.² I suggest, however, that, although thesis (C), “the phenomenalist version,” is false, the two theses, (A) and (B), which constitute the doctrine of the given are true.

The doctrine is not merely the consequence of a metaphor. We are led to it when we attempt to answer certain questions about *justification* – our justification for supposing, in connection with any one of the things that we know to be true, that it is something that we know to be true.

2. To the question “What justification do I have for thinking I know that *a* is true?” one may reply: “I know that *b* is true, and if I know that *b* is true then I also know that *a* is true.” And to the question “What justification do I have for thinking I know that *b* is true?” one may reply: “I know that *c* is true, and if I know that *c* is true then I also know that *b* is true.” Are we thus led, sooner or later, to something *n* of which one may say: “What justifies me in thinking I know that *n* is true is simply the fact that *n* is true.” If there is such an *n*, then the belief or statement that *n* is true may be thought of either as a belief or statement which “justifies itself” or as a belief or statement which is itself “neither justified nor unjustified.” The distinction – unlike that between a Prime Mover which moves itself and a Prime Mover which is neither in motion nor at rest – is largely a verbal one; the essential thing, if there is such an *n*, is that it provides a stopping place in the process, or dialectic, of justification.

We may now re-express, somewhat less metaphorically, the two theses which I have called the “doctrine of the given.” The first thesis, that our knowledge is an edifice or structure having its own foundation, becomes (A) “every statement, which we are justified in thinking that we know, is justified in part by some statement which justifies itself.” The second thesis, that there are appear-

ances (“the given”) at the foundation of our knowledge, becomes (B) “there are statements about appearances which thus justify themselves.” (The third thesis – the “phenomenalist version” of the doctrine of the given – becomes (C) “there are no self-justifying statements which are not statements about appearances.”)

Let us now turn to the first of the two theses constituting the doctrine of the given.

3. “Every justified statement is justified in part by some statement which justifies itself.” Could it be that the question which this thesis is supposed to answer is a question which arises only because of some mistaken assumption? If not, what are the alternative ways of answering it? And did any of the philosophers with whom we are concerned actually accept any of these alternatives? The first two questions are less difficult to answer than the third.

There are the following points of view to be considered, each of which *seems* to have been taken by some of the philosophers in the period of our survey.

- (1) One may believe that the questions about justification which give rise to our problem are based upon false assumptions and hence that they *should not be asked* at all.
- (2) One may believe that no statement or claim is justified unless it is justified, at least in part, by some other justified statement or claim which it does not justify; this belief may suggest that one should continue the process of justifying *ad indefinitum*, justifying each claim by reference to some additional claim.
- (3) One may believe that no statement or claim *a* is justified unless it is justified by some other justified statement or claim *b*, and that *b* is not justified unless it in turn is justified by *a*; this would suggest that the process of justifying is, or should be, *circular*.
- (4) One may believe that there are some particular claims *n* at which the process of justifying should stop, and one may then hold of any such claim *n* either: (a) *n* is justified by something – viz., *experience* or *observation* – which is not itself a claim and which therefore cannot be said itself either to be justified or unjustified; (b) *n* is itself *unjustified*; (c) *n* justifies itself; or (d) *n* is *neither justified nor unjustified*.

These possibilities, I think, exhaust the significant points of view; let us now consider them in turn.

4. "The questions about justification which give rise to the problem are based upon false assumptions and therefore should not be asked at all."

The questions are *not* based upon false assumptions; but most of the philosophers who discussed the questions put them in such a misleading way that one is very easily misled into supposing that they *are* based upon false assumptions.

Many philosophers, following Descartes, Russell, and Husserl, formulated the questions about justification by means of such terms as "doubt," "certainty," and "incorrigibility," and they used, or misused, these terms in such a way that, when their questions were taken in the way in which one would ordinarily take them, they could be shown to be based upon false assumptions. One may note, for example, that the statement "There is a clock on the mantelpiece" is not self-justifying – for to the question "What is your justification for thinking you know that there is a clock on the mantelpiece?" the proper reply would be to make some other statement (e.g., "I saw it there this morning and no one would have taken it away") – and one may then go on to ask "But are there any statements which can be said to justify themselves?" If we express these facts, as many philosophers did, by saying that the statement "There is a clock on the mantelpiece" is one which is not "certain," or one which may be "doubted," and if we then go on to ask "Does this doubtful statement rest upon other statements which are certain and incorrigible?" then we are using terms in an extraordinarily misleading way. The question "Does this doubtful statement rest upon statements which are certain and incorrigible?" – if taken as one would ordinarily take it – does rest upon a false assumption, for (we may assume) the statement that there is a clock on the mantelpiece is one which is not doubtful at all.

John Dewey, and some of the philosophers whose views were very similar to his, tended to suppose, mistakenly, that the philosophers who asked themselves "What justification do I have for thinking I know this?" were asking the quite different question "What more can I do to verify or confirm that this is so?" and they rejected answers to the first question on the ground that they were unsatisfactory answers to the second.³ Philosophers influenced by Wittgenstein tended to

suppose, also mistakenly, but quite understandably, that the question "What justification do I have for thinking I know this?" contains an implicit challenge and presupposes that one does not have the knowledge concerned. They then pointed out, correctly, that in most of the cases where the question was raised (e.g., "What justifies me in thinking I know that this is a table?") there is no ground for challenging the claim to knowledge and that questions presupposing that the claim is false should not arise. But the question "What justifies me in thinking I know that this is a table?" does not challenge the claim to know that this is a table, much less presuppose that the claim is false.

The "critique of cogency," as Lewis described this concern of epistemology, presupposes that we *are* justified in thinking we know most of the things that we do think we know, and what it seeks to elicit is the nature of this justification. The enterprise is like that of ethics, logic, and aesthetics:

The nature of the good can be learned from experience only if the content of experience be first classified into good and bad, or grades of better and worse. Such classification or grading already involves the legislative application of the same principle which is sought. In logic, principles can be elicited by generalization from examples only if cases of valid reasoning have first been segregated by some criterion. In esthetics, the laws of the beautiful may be derived from experience only if the criteria of beauty have first been correctly applied.⁴

When Aristotle considered an invalid mood of the syllogism and asked himself "What is wrong with this?" he was not suggesting to himself that perhaps nothing was wrong; he presupposed that the mood *was* invalid, just as he presupposed that others were not, and he attempted, successfully, to formulate criteria which would enable us to distinguish the two types of mood.

When we have answered the question "What justification do I have for thinking I know this?" what we learn, as Socrates taught, is something about ourselves. We learn, of course, what the justification happens to be for the particular claim with which the question is concerned. But we also learn, more generally, what the criteria are, if any, in terms of which we believe ourselves justified in counting one thing as an instance of knowing and another thing not. The truth which

the philosopher seeks, when he asks about justification, is "already implicit in the mind which seeks it, and needs only to be elicited and brought to clear expression."⁵

Let us turn, then, to the other approaches to the problem of "the given."

5. "No statement or claim would be justified unless it were justified, at least in part, by some other justified claim or statement which it does not justify."

This regressive principle might be suggested by the figure of the building and its supports: no stage supports another unless it is itself supported by some other stage beneath it – a truth which holds not only of the upper portions of the building but also of what we call its foundation. And the principle follows if, as some of the philosophers in the tradition of logical empiricism seemed to believe, we should combine a frequency theory of probability with a probability theory of justification.

In *Experience and Prediction* (U. of Chicago, 1938) and in other writings, Hans Reichenbach defended a "probability theory of knowledge" which seemed to involve the following contentions:

- (1) To justify accepting a statement, it is necessary to show that the statement is probable.
- (2) To say of a statement that it is probable is to say something about statistical frequencies. Somewhat more accurately, a statement of the form "It is *probable* that any particular *a* is a *b*" may be explicated as saying "Most *a*'s are *b*'s." Or, still more accurately, to say "The probability is *n* that a particular *a* is a *b*" is to say "The limit of the relative frequency with which the property of being a *b* occurs in the class of things having the property *a* is *n*."
- (3) Hence, by (2), to show that a proposition is probable it is necessary to show that a certain statistical frequency obtains; and, by (1), to show that a certain statistical frequency obtains it is necessary to show that it is probable that the statistical frequency obtains; and therefore, by (2), to show that it is probable that a certain statistical frequency obtains, it is necessary to show that a certain frequency of frequencies obtains. . . .
- (4) And therefore "there is no Archimedean point of absolute certainty left to which to attach our knowledge of the world; all we

have is an elastic net of probability connections floating in open space" (p. 192).

This reasoning suggests that an infinite number of steps must be taken in order to justify acceptance of any statement. For, according to the reasoning, we cannot determine the probability of one statement until we have determined that of a second, and we cannot determine that of the second until we have determined that of a third, and so on. Reichenbach does not leave the matter here, however. He suggests that there is a way of "descending" from this "open space" of probability connections, but, if I am not mistaken, we can make the descent only by letting go of the concept of justification.

He says that, if we are to avoid the regress of probabilities of probabilities of probabilities . . . we must be willing at some point merely to make a guess; "there will always be some blind posits on which the whole concatenation is based" (p. 367). The view that knowledge is to be identified with certainty and that probable knowledge must be "imbedded in a framework of certainty" is "a remnant of rationalism. An empiricist theory of probability can be constructed only if we are willing to regard knowledge as a system of posits."⁶

But if we begin by assuming, as we do, that there is a distinction between knowledge, on the one hand, and a lucky guess, on the other, then we must reject at least one of the premises of any argument purporting to demonstrate that knowledge is a system of "blind posits." The unacceptable conclusion of Reichenbach's argument may be so construed as to follow from premises (1) and (2); and premise (2) may be accepted as a kind of definition (though there are many who believe that this definition is not adequate to all of the uses of the term "probable" in science and everyday life). Premise (1), therefore, is the one we should reject, and there are good reasons, I think, for rejecting (1), the thesis that "to justify accepting a proposition it is necessary to show that the proposition is probable." In fairness to Reichenbach, it should be added that he never explicitly affirms premise (1); but some such premise is essential to his argument.

6. "No statement or claim *a* would be justified unless it were justified by some other justified statement or claim *b* which would not be justified unless it were justified in turn by *a*."

The "coherence theory of truth," to which some philosophers committed themselves, is

sometimes taken to imply that justification may thus be circular; I believe, however, that the theory does not have this implication. It does define "truth" as a kind of systematic consistency of beliefs or propositions. The truth of a proposition is said to consist, not in the fact that the proposition "corresponds" with something which is not itself a proposition, but in the fact that it fits consistently into a certain more general system of propositions. This view may even be suggested by the figure of the building and its foundations. There is no difference in principle between the way in which the upper stories are supported by the lower, and that in which the cellar is supported by the earth just below it, or the way in which that stratum of earth is supported by various substrata farther below; a good building appears to be a part of the terrain on which it stands and a good system of propositions is a part of the wider system which gives it its truth. But these metaphors do not solve philosophical problems.

The coherence theory did in fact appeal to something other than logical consistency; its proponents conceded that a system of false propositions may be internally consistent and hence that logical consistency alone is no guarantee of truth. Brand Blanshard, who defended the coherence theory in *The Nature of Thought*, said that a proposition is true provided it is a member of an internally consistent system of propositions and *provided further* this system is "the system in which everything real and possible is coherently included."⁷ In one phase of the development of "logical empiricism" its proponents seem to have held a similar view: a proposition – or, in this case, a statement – is true provided it is a member of an internally consistent system of statements and *provided further* this system is "the system which is actually adopted by mankind, and especially by the scientists in our culture circle."⁸

A theory of truth is not, as such, a theory of justification. To say that a proposition is true is not to say that we are justified in accepting it as true, and to say that we are justified in accepting it as true is not to say that it is true. Whatever merits the coherence theory may have as an answer to certain questions about truth, it throws no light upon our present epistemological question. If we accept the coherence theory, we may still ask, concerning any proposition *a* which we think we know to be true, "What is my justification for thinking I know that *a* is a member of the system of propositions in which everything real and poss-

ible is coherently included, or that *a* is a member of the system of propositions which is actually adopted by mankind and by the scientists of our culture circle?" And when we ask such a question, we are confronted, once again, with our original alternatives.

7. If our questions about justification do have a proper stopping place, then, as I have said, there are still four significant possibilities to consider. We may stop with some particular claim and say of it that either:

- (a) it is justified by something – by experience, or by observation – which is not itself a claim and which, therefore, cannot be said either to be justified or to be unjustified;
- (b) it is justified by some claim which refers to our experience or observation, and the claim referring to our experience or observation has *no* justification;
- (c) it justifies itself; or
- (d) it is itself neither justified nor unjustified.

The first of these alternatives leads readily to the second, and the second to the third or to the fourth. The third and the fourth – which differ only verbally, I think – involve the doctrine of "the given."

Carnap wrote, in 1936, that the procedure of scientific testing involves two operations: the "confrontation of a statement with observation" and the "confrontation of a statement with previously accepted statements." He suggested that those logical empiricists who were attracted to the coherence theory of truth tended to lose sight of the first of these operations – the confrontation of a statement with observation. He proposed a way of formulating simple "acceptance rules" for such confrontation and he seemed to believe that, merely by applying such rules, we could avoid the epistemological questions with which the adherents of "the given" had become involved.

Carnap said this about his acceptance rules: "If no foreign language or introduction of new terms is involved, the rules are trivial. For example: 'If one is hungry, the statement "I am hungry" may be accepted'; or: 'If one sees a key one may accept the statement "there lies a key."'"⁹ As we shall note later, the first of these rules differs in an important way from the second. Confining ourselves for the moment to rules of the second sort – "If one sees a key one may accept the statement 'there lies a key'" – let us ask ourselves whether

the appeal to such rules enables us to solve our problem of the stopping place.

When we have made the statement "There lies a key," we can, of course, raise the question "What is my justification for thinking I know, or for believing, that there lies a key?" The answer would be "I see the key." We cannot ask "What is my justification for seeing a key?" But we *can* ask "What is my justification for thinking that it is a key that I see?" and, if we *do* see that the thing is a key, the question will have an answer. The answer might be "I see that it's shaped like a key and that it's in the lock, and I remember that a key is usually here." The possibility of this question, and its answer, indicates that we cannot stop our questions about justification merely by appealing to observation or experience. For, of the statement "I observe that that is an A," we can ask, and answer, the question "What is my justification for thinking that I observe that there is an A?"

It is relevant to note, moreover, that there may be conditions under which seeing a key does *not* justify one in accepting the statement "There is a key" or in believing that one sees a key. If the key were so disguised or concealed that the man who saw it did not recognize it to be a key, then he might not be justified in accepting the statement "There is a key." Just as, if Mr. Jones unknown to anyone but himself is a thief, then the people who see him may be said to see a thief – but none of those who thus sees a thief is justified in accepting the statement "There is a thief."¹⁰

Some of the writings of logical empiricists suggest that, although some statements may be justified by reference to other statements, those statements which involve "confrontation with observation" are not justified at all. C. G. Hempel, for example, wrote that "the acknowledgement of an experiential statements as true is psychologically motivated by certain experiences; but within the system of statements which express scientific knowledge or one's beliefs at a given time, they function in the manner of postulates for which no grounds are offered."¹¹ Hempel conceded, however, that this use of the term "postulate" is misleading and he added the following note of clarification: "When an experiential sentence is accepted 'on the basis of direct experiential evidence,' it is indeed not asserted arbitrarily; but to describe the evidence in question would simply mean to repeat the experiential statement itself. Hence, in the context of cognitive justification,

the statement functions in the manner of a primitive sentence."¹²

When we reach a statement having the property just referred to – an experiential statement such that to describe its evidence "would simply mean to repeat the experiential statement itself" – we have reached a proper stopping place in the process of justification.

8. We are thus led to the concept of a belief, statement, claim, proposition, or hypothesis, which justifies itself. To be clear about the concept, let us note the way in which we would justify the statement that we have a certain belief. It is essential, of course, that we distinguish justifying the statement *that* we have a certain belief from justifying the belief itself.

Suppose, then, a man is led to say "I believe that Socrates is mortal" and we ask him "What is your justification for thinking that you believe, or for thinking that you know that you believe, that Socrates is mortal?" To this strange question, the only appropriate reply would be "My justification for thinking I believe, or for thinking that I know that I believe, that Socrates is mortal is simply the fact that I *do* believe that Socrates is mortal." One justifies the statement simply by reiterating it; the statement's justification is what the statement says. Here, then, we have a case which satisfies Hempel's remark quoted above; we describe the evidence for a statement merely by repeating the statement. We could say, as C. J. Ducasse did, that "the occurrence of belief is its own evidence."¹³

Normally, as I have suggested, one cannot justify a statement merely by reiterating it. To the question "What justification do you have for thinking you know that there can be no life on the moon?" it would be inappropriate, and impertinent, to reply by saying simply "There *can* be no life on the moon," thus reiterating the fact at issue. An appropriate answer would be one referring to certain *other* facts – for example, the fact that we know there is insufficient oxygen on the moon to support any kind of life. But to the question "What is your justification for thinking you know that you believe so and so?" there is nothing to say other than "I *do* believe so and so."

We may say, then, that there are some statements which are self-justifying, or which justify themselves. And we may say, analogously, that there are certain beliefs, claims, propositions, or hypotheses which are self-justifying, or which

justify themselves. A statement, belief, claim, proposition, or hypothesis may be said to be self-justifying for a person, if the person's justification for thinking he knows it to be true is simply the fact that it *is* true.

Paradoxically, these things I have described by saying that they "justify themselves" may *also* be described by saying they are "neither justified nor unjustified." The two modes of description are two different ways of saying the same thing.

If we are sensitive to ordinary usage, we may note that the expression "I believe that I believe" is ordinarily used, not to refer to a second-order belief about the speaker's own beliefs, but to indicate that the speaker has not yet made up his mind. "I believe that I believe that Johnson is a good president" might properly be taken to indicate that, if the speaker *does* believe that Johnson is a good president, he is not yet firm in that belief. Hence there is a temptation to infer that, if we say of a man who is firm in his belief that Socrates is mortal, that he is "justified in believing that he believes that Socrates is mortal," our statement "makes no sense." And there is also a temptation to go on and say that it "makes no sense" even to say of such a man, that his *statement* "I believe that Socrates is mortal" is one which is "justified" for him.¹⁴ After all, what would it mean to say of a man's statement about his own belief, that he is *not* justified in accepting it?¹⁵

The questions about what does or does not "make any sense" need not, however, be argued. We *may* say, if we prefer, that the statements about the beliefs in question are "neither justified nor unjustified." Whatever mode of description we use, the essential points are two. First, we may appeal to such statements in the process of justifying some *other* statement or belief. If they *have* no justification they may yet *be* a justification – for something other than themselves. ("What justifies me in thinking that he and I are not likely to agree? The fact that I believe that Socrates is mortal and he does not.") Second, the making of such a statement does provide what I have been calling a "stopping place" in the dialectic of justification; but now, instead of signaling the stopping place by reiterating the questioned statement, we do it by saying that the question of its justification is one which "should not arise."

It does not matter, then, whether we speak of certain statements which "justify themselves" or of certain statements which are "neither justified nor unjustified," for in either case we will be

referring to the same set of statements. I shall continue to use the former phrase.

There are, then, statements about one's own beliefs ("I believe that Socrates is mortal") – and for statements about many other psychological attitudes – which are self-justifying. "What justifies me in believing, or in thinking I know, that I *hope* to come tomorrow? Simply the fact that I *do* hope to come tomorrow." Thinking, desiring, wondering, loving, hating, and other such attitudes are similar. Some, but by no means all, of the statements we can make about such attitudes, when the attitudes are our own, are self-justifying – as are statements containing such phrases as "I think I remember" or "I seem to remember" (as distinguished from "I remember"), and "I think that I see" and "I think that I perceive" (as distinguished from "I see" and "I perceive"). Thus, of the two examples which Carnap introduced in connection with his "acceptance rules" discussed above, viz., "I am hungry" and "I see a key," we may say that the first is self-justifying and the second not.

The "doctrine of the given," it will be recalled, tells us (A) that every justified statement, about what we think we know, is justified in part by some statement which justifies itself and (B) that there are statements about appearances which thus justify themselves. The "phenomenalistic version" of the theory adds (C) that statements about appearances are the *only* statements which justify themselves. What we have been saying is that the first thesis, (A), of the doctrine of the given is true and that the "phenomenalistic version," (C), is false; let us turn now to thesis (B).

9. In addition to the self-justifying statements about psychological attitudes, are there self-justifying statements about "appearances"? Now we encounter difficulties involving the word "appearance" and its cognates.

Sometimes such words as "appears," "looks," and "seems" are used to convey what one might also convey by such terms as "believe." For example, if I say "It appears to me that General de Gaulle was successful," or "General de Gaulle seems to have been successful," I am likely to mean only that I believe, or incline to believe, that he has been successful; the words "appears" and "seems" serve as useful hedges, giving me an out, should I find out later that de Gaulle was not successful. When "appear"-words are used in this way, the statements in which they occur add

nothing significant to the class of "self-justifying" statements we have just provided. Philosophers have traditionally assumed, however, that such terms as "appear" may also be used in a quite different way. If this assumption is correct, as I believe it is, then this additional use does lead us to another type of self-justifying statement.

The philosophers who exposed the confusions to which the substantival expression "appearance" gave rise were sometimes inclined to forget, I think, that things do appear to us in various ways.¹⁶ We can alter the appearance of anything we like merely by doing something which will affect our sense organs or the conditions of observation. One of the important epistemological questions about appearances is "Are there self-justifying statements about the ways in which things appear?"

Augustine, refuting the skeptics of the late Platonic Academy, wrote:

I do not see how the Academician can refute him who says: I know that this appears white to me, I know that my hearing is delighted with this, I know this has an agreeable odor, I know this tastes sweet to me, I know that this feels cold to me When a person tastes something, he can honestly swear that he knows it is sweet to his palate or the contrary, and that no trickery of the Greeks can dispossess him of that knowledge.¹⁷

Suppose, now, one were to ask "What justification do you have for believing, or thinking you know, that this appears white to you, or that that tastes bitter to you?" Here, too, we can only reiterate the statement: "What justifies me in believing, or in thinking I know, that this appears white to me and that that tastes bitter to me is the fact that this *does* appear white to me and that *does* taste bitter."

An advantage of the misleading substantive "appearance," as distinguished from the verb "appears," is that the former may be applied to those sensuous experiences which, though capable of being appearances of things, are actually not appearances of anything. Feelings, imagery, and the sensuous content of dreams and hallucination are very much like the appearances of things and they are such that, under some circumstances, they could be appearances of things. But if we do not wish to say that they are experiences wherein some external physical thing *appears* to us, we must use some expression other than "appear." For

"appear," in its active voice, requires a grammatical subject and thus requires a term which refers, not merely to a way of appearing, but also to something *which* appears.

But we may avoid *both* the objective "*Something appears blue to me,*" and the substantival "I sense a blue *appearance.*" We may use another verb, say "sense," in a technical way, as many philosophers did, and equate it in meaning with the passive voice of "appear," thus saying simply "I *sense* blue," or the like. Or better still, it seems to me, and at the expense only of a little awkwardness, we can use "appear" in its passive voice and say "I am *appeared to* blue."

Summing up, in our new vocabulary, we may say that the philosophers who talked of the "empirically given" were referring, not to "self-justifying" statements and beliefs generally, but only to those pertaining to certain "ways of being appeared to." And the philosophers who objected to the doctrine of the given, or some of them, argued, that no statement about "a way of being appeared to" can be "self-justifying."

10. Why would one suppose that "This appears white" (or, more exactly, "I am now appeared white to") is not self-justifying? The most convincing argument was this: If I say "This appears white," then, as Reichenbach put it, I am making a "comparison between a present object and a formerly seen object."¹⁸ What I am saying *could* have been expressed by "The present way of appearing is the way in which white objects, or objects which I believe to be white, ordinarily appear." And this new statement, clearly, is not self-justifying; to justify it, as Reichenbach intimated, I must go on and say something further — something about the way in which I remember white objects to have appeared.

"Appears white" *may* thus be used to abbreviate "appears the way in which white things normally appear." Or "white thing," on the other hand, *may* be used to abbreviate "thing having the color of things which ordinarily appear white." The phrase "appear white" as it is used in the second quoted expression cannot be spelled out in the manner of the first; for the point of the second can hardly be put by saying that "white thing" may be used to abbreviate "thing having the color of things which ordinarily appear the way in which *white things* normally appear." In the second expression, the point of "appears white" is not to *compare* a way of appearing with something else; the point is to say

something about the way of appearing itself. It is in terms of this second sense of “appears white” – that in which one may say significantly and without redundancy “Things that are white may normally be expected to appear white” – that we are to interpret the quotation from Augustine above. And, more generally, when it was said that “appear”-statements constitute the foundation of the edifice of knowledge, it was not intended that the “appear”-statements be interpreted as statements asserting a comparison between a present object and any other object or set of objects.

The question now becomes “Can we formulate any significant ‘appear’-statements *without* thus comparing the way in which some object appears with the way in which some other object appears, or with the way in which the object in question has appeared at some other time? Can we interpret ‘This appears white’ in such a way that it may be understood to refer to a present way of appearing *without* relating that way of appearing to any other object?” In *Experience and Prediction*, Reichenbach defended his own view (and that of a good many others) in this way:

The objection may be raised that a comparison with formerly seen physical objects should be avoided, and that a basic statement is to concern the present fact only, as it is. But such a reduction would make the basic statement empty. Its content is just that there is a similarity between the present object and one formerly seen; it is by means of this relation that the present object is described. Otherwise the basic statement would consist in attaching an individual symbol, say a number, to the present object; but the introduction of such a symbol would help us in no way, since we could not make use of it to construct a comparison with other things. Only in attaching the same symbols to different objects, do we arrive at the possibility of constructing relations between the objects. (pp. 176–7)

It is true that, if an “appear”-statement is to be used successfully in communication, it must assert some comparison of objects. Clearly, if I wish *you* to know the way things are now appearing to me, I must relate these ways of appearing to something that is familiar to you. But our present question is not “Can you understand me if I predicate something of the way in which something now appears to me without relating that way of appearing to

something that is familiar to you?” The question is, more simply, “Can I predicate anything of the way in which something now appears to me without thereby comparing that way of appearing with something else?” From the fact that the first of these two questions must be answered in the negative it does not follow that the second must also be answered in the negative.¹⁹

The issue is not one about communication, nor is it, strictly speaking, an issue about language; it concerns, rather, the nature of thought itself. Common to both “pragmatism” and “idealism,” as traditions in American philosophy, is the view that to *think* about a thing, or to *interpret* or *conceptualize* it, and hence to have a *belief* about it, is essentially to relate the thing to *other* things, actual or possible, and therefore to “refer beyond it.” It is this view – and not any view about language or communication – that we must oppose if we are to say of some statements about appearing, or of any other statements, that they “justify themselves.”

To think about the way in which something is now appearing, according to the view in question, is to relate that way of appearing to something else, possibly to certain future experiences, possibly to the way in which things of a certain sort may be commonly expected to appear. According to the “conceptualistic pragmatism” of C. I. Lewis’s *Mind and the World-Order* (1929), we grasp the present experience, any present way of appearing, only to the extent to which we relate it to some future experience.²⁰ According to one interpretation of John Dewey’s “instrumentalistic” version of pragmatism, the present experience may be used to present or disclose something else but it does not present or disclose itself. And according to the idealistic view defended in Brand Blanshard’s *The Nature of Thought*, we grasp our present experience only to the extent that we are able to include it in the one “intelligible system of universals” (vol. I, p. 632).

This theory of reference, it should be noted, applies not only to statements and beliefs about “ways of being appeared to” but also to those other statements and beliefs which I have called “self-justifying.” If “This appears white,” or “I am appeared white to,” compares the present experience with something else, and thus depends for its justification upon what we are justified in believing about the something else, then so, too, does “I believe that Socrates is mortal” and “I hope that the peace will continue.” This general conception of thought, therefore, would seem to imply that no

belief or statement can be said to justify itself. But according to what we have been saying, if there is no belief or statement which justifies itself, then it is problematic whether any belief or statement is justified at all. And therefore, as we might expect, this conception of thought and reference has been associated with skepticism.

Blanshard conceded that his theory of thought "does involve a degree of scepticism regarding our present knowledge and probably all future knowledge. In all likelihood there will never be a proposition of which we can say, 'This that I am asserting, with precisely the meaning I now attach to it, is absolutely true.'"²¹ On Dewey's theory, or on one common interpretation of Dewey's theory, it is problematic whether anyone can now be said to *know* that Mr Jones is working in his garden. A. O. Lovejoy is reported to have said that, for Dewey, "I am about to have known" is as close as we ever get to "I know."²² C. I. Lewis, in his *An Analysis of Knowledge and Valuation* (Open Court, 1946) conceded in effect that the conception of thought suggested by his earlier *Mind and the World-Order* does lead to a kind of skepticism; according to the later work there are "apprehensions of the given" (cf. *An Analysis*, pp. 182-3) – and thus beliefs which justify themselves.

What is the plausibility of a theory of thought and reference which seems to imply that no one knows anything?

Perhaps it is correct to say that when we think about a thing we think about it as having certain properties. But why should one go on to say that to think about a thing must always involve thinking about some *other* thing as well? Does thinking about the other thing then involve thinking about some third thing? Or can we think about one thing in relation to a second thing without thereby thinking of a third thing? And if we can, then why can we not think of one thing – of one thing as having certain properties – without thereby relating it to another thing?

The linguistic analogue of this view of thought is similar. Why should one suppose – as Reichenbach supposed in the passage cited above and as many others have also supposed – that to *refer* to a thing, in this instance to refer to a way of appearing, is necessarily to relate the thing to some *other* thing?

Some philosophers seem to have been led to such a view of reference as a result of such considerations as the following: We have imagined a man saying, in agreement with Augustine, "It just

does appear white – and that is the end of the matter." Let us consider now the possible reply "That it is not the end of the matter. You are making certain assumptions about the language you are using; you are assuming, for example, that you are using the word 'white,' or the phrase 'appears white,' in the way in which you have formerly used it, or in the way in which it is ordinarily used, or in the way in which it would ordinarily be understood. And if you state your justification for this assumption, you *will* refer to certain other things – to yourself and to other people, to the word 'white,' or to the phrase 'appears white,' and to what the word or phrase has referred to or might refer to on other occasions. And therefore, when you say 'This appears white' you are saying something, not only about your present experience, but also about all of these other things as well."

The conclusion of this argument – the part that follows the "therefore" – does not follow from the premises. In supposing that the argument is valid, one fails to distinguish between (1) *what* it is that a man means to say when he uses certain words and (2) his assumptions concerning the adequacy of these words for *expressing* what it is that he means to say; one supposes, mistakenly, that what justifies (2) must be included in what justifies (1). A Frenchman, not yet sure of his English, may utter the words "There are apples in the basket," intending thereby to express his belief that there are potatoes in the basket. If we show him that he has used the word "apples" incorrectly, and hence that he is mistaken in his assumptions about the ways in which English speaking people use and understand the word "apples," we have not shown him anything relevant to his *belief* that there are apples in the basket.

Logicians now take care to distinguish between the *use* and *mention* of language (e.g., the English word "Socrates" is mentioned in the sentence "Socrates' has eight letters" and is used but not mentioned, in "Socrates is a Greek.")²³ As we shall have occasion to note further, the distinction has not always been observed in writings on epistemology.

11. If we decide, then, that there is a class of beliefs or statements which are "self-justifying," and that this class is limited to certain beliefs or statements about our own psychological states and about the ways in which we are "appeared to," we may be tempted to return to the figure of the edifice: our

knowledge of the world is a structure supported entirely by a foundation of such self-justifying statements or beliefs. We should recall, however, that the answers to our original Socratic questions had *two* parts. When asked "What is your justification for thinking that you know *a*?" one may reply "I am justified in thinking I know *a*, because (1) I know *b* and (2) if I know *b* then I know *a*." We considered our justification for the *first* part of this answer, saying "I am justified in thinking I know *b*, because (1) I know *c* and (2) if I know *c* then I know *b*." And then we considered our justification for the first part of the second answer, and continued in this fashion until we reached the point of self-justification. In thus moving toward "the given," we accumulated, step by step, a backlog of claims that we did not attempt to justify – those claims constituting the *second* part of each of our answers. Hence our original claim – "I know that *a* is true" – does not rest upon "the given" alone; it also rests upon all of those other claims that we made *en route*. And it is not justified unless these other claims are justified.

A consideration of these other claims will lead us, I think, to at least three additional types of "stopping place," which are concerned, respectively, with memory, perception, and what Kant called the *a priori*. Here I shall comment briefly on the first two.

It is difficult to think of any claim to empirical knowledge, other than the self-justifying statements we have just considered, which does not to some extent rest upon an appeal to memory. But the appeal to memory – "I remember that *A* occurred" – is not self-justifying. One may ask "And what is your justification for thinking that you remember that *A* occurred?" and the question will have an answer – even if the answer is only the self-justifying "I think that I remember that *A* occurred." The statement "I remember that *A* occurred" does, of course, imply "*A* occurred"; but "I think that I remember that *A* occurred" does not imply "*A* occurred" and hence does not imply "I remember that *A* occurred." For we can remember occasions – at least we think we can remember them – when we learned, concerning some event we had thought we remembered, that the event had not occurred at all, and consequently that we had not really remembered it. When we thus find that one memory conflicts with another, or, more accurately, when we thus find that one thing that we think we remember conflicts with another thing that we think we remember, we may

correct one or the other by making further inquiry; but the results of any such inquiry will always be justified in part by other memories, or by other things that we think that we remember. How then are we to choose between what seem to be conflicting memories? Under what conditions does "I think that I remember that *A* occurred" serve to justify "I remember that *A* occurred"?

The problem is one of formulating a rule of evidence – a rule specifying the conditions under which statements about what we think we remember can justify statements about what we do remember. A possible solution, in very general terms, is "When we think that we remember, then we are justified in believing that we do remember, provided that what we think we remember does not conflict with anything else that we think we remember; when what we think we remember does conflict with something else we think we remember, then, of the two conflicting memories (more accurately, ostensible memories) the one that is justified is the one that fits in better with the other things that we think we remember." Ledger Wood made the latter point by saying that the justified memory is the one which "coheres with the system of related memories"; C. I. Lewis used "congruence" instead of "coherence."²⁴ But we cannot say precisely what is meant by "fitting in," "coherence," or "congruence" until certain controversial questions of confirmation theory and the logic of probability have been answered. And it may be that the rule of evidence is too liberal; perhaps we should say, for example, that when two ostensible memories conflict neither one of them is justified. But these are questions which have not yet been satisfactorily answered.

If we substitute "perceive" for "remember" in the foregoing, we can formulate a similar set of problems about perception; these problems, too, must await solution.²⁵

The problems involved in formulating such rules of evidence, and in determining the validity of these rules, do not differ in any significant way from those which arise in connection with the formulation, and validity, of the rules of logic. Nor do they differ from the problems posed by the moral and religious "cognitivists" (the "non-intuitionistic cognitivists") that I have referred to elsewhere. The status of ostensible memories and perceptions, with respect to that experience which is their "source," is essentially like that which such "cognitivists" claim for judgments having an ethical or theological subject matter. Unfortunately, it

is also like that which other "enthusiasts" claim for still other types of subject matter.

12. What, then, is the status of the doctrine of "the given" – of the "myth of the given"? In my opinion, the doctrine is correct in saying that there are some beliefs or statements which are "self-justifying" and that among such beliefs and statements are some which concern appearances or "ways of being appeared to;" but the "phenomenalistic version" of the doctrine is mistaken in implying that our knowledge may be thought of as an edifice which is supported by appearances alone.²⁶ The cognitive significance of "the empirically given" was correctly described – in a vocabulary rather different from that which I have been using – by John Dewey:

The alleged primacy of sensory meanings is mythical. They are primary only in logical status; they are primary as tests and confirmation

of inferences concerning matters of fact, not as historic originals. For, while it is not usually needful to carry the check or test of theoretical calculations to the point of irreducible *sensa*, colors, sounds, etc., these *sensa* form a limit approached in careful analytic certifications, and upon critical occasions it is necessary to touch the limit... *Sensa* are the class of irreducible meanings which are employed in verifying and correcting other meanings. We actually set out with much coarser and more inclusive meanings and not till we have met with failure from their use do we even set out to discover those ultimate and harder meanings which are sensory in character.²⁷

The Socratic questions leading to the concept of "the given" also lead to the concept of "rules of evidence." Unfortunately some of the philosophers who stressed the importance of the former concept tended to overlook that of the latter.

Notes

- 1 Philosophers in other traditions also noted these confusions. See, for example, John Wild, "The Concept of the Given in Contemporary Philosophy," *Philosophy and Phenomenological Research* I (1940), pp. 70–82.
- 2 The expression "myth of the given" was used by Wilfrid Sellars in "Empiricism and the Philosophy of Mind," in Herbert Feigl and Michael Scriven (eds), *Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minnesota Studies in the Philosophy of Science, vol. I (U. of Minn., 1956), pp. 253–329.
- 3 Dewey also said that, instead of trying to provide "Foundations for Knowledge," the philosopher should apply "what is known to intelligent conduct of the affairs of human life" to "the problems of men." John Dewey, *Problems of Men* (Philosophical, 1946), pp. 6–7.
- 4 C. I. Lewis, *Mind and the World-Order* (Scribner, 1929), p. 29.
- 5 *Ibid.*, p. 19. Cf. Hans Reichenbach, *Experience and Prediction* (University of Chicago, 1938), p. 6; C. J. Ducasse, "Some Observations Concerning the Nature of Probability," *Journal of Philosophy* XXXVIII (1941), esp. pp. 400–1.
- 6 Hans Reichenbach, "Are Phenomenal Reports Absolutely Certain?" *Philosophical Review* LXI (1952), pp. 147–59; the quotation is from p. 150.
- 7 Brand Blanshard, *The Nature of Thought*, vol. II (Macmillan, 1940), p. 276.
- 8 C. G. Hempel, "On the Logical Positivists' Theory of Truth," *Analysis* II (1935), pp. 49–59; the quotation is from p. 57.
- 9 Rudolf Carnap, "Truth and Confirmation," in Herbert Feigl and W. S. Sellars (eds), *Readings in Philosophical Analysis* (Appleton, 1949), p. 125. The portions of the article quoted above first appeared in "Wahrheit und Bewahrung," *Actes du congrès international de philosophie scientifique*, IV (Paris; 1936), pp. 18–23.
- 10 Cf. Nelson Goodman, *The Structure of Appearance* (Harvard, 1951), p. 104. If Goodman's book, incidentally, is not discussed in this collection of essays, the fault is with our conventional classification of philosophical disciplines. The book, which is concerned with an area falling between logic and metaphysics, is one of the most important philosophical works written by an American during the period being surveyed.
- 11 C. G. Hempel, "Some Theses on Empirical Certainty," *Review of Metaphysics* V (1952), pp. 621–9; the quotation is from p. 621.
- 12 *Ibid.*, p. 628. Hempel's remarks were made in an "Exploration" in which he set forth several theses about "empirical certainty" and then replied to objections by Paul Weiss, Roderick Firth, Wilfrid Sellars, and myself.
- 13 C. J. Ducasse, "Propositions, Truth, and the Ultimate Criterion of Truth," *Philosophy and*

- Phenomenological Research* IV (1939), pp. 317–40; the quotation is from p. 339.
- 14 Cf. Norman Malcolm, "Knowledge of Other Minds," *Journal of Philosophy* LV (1958), pp. 969–78. Reprinted in Malcolm, *Knowledge and Certainty: Essays and Lectures* (Prentice-Hall, 1963).
 - 15 The principle behind this way of looking at the matter is defended in detail by Max Black in *Language and Philosophy* (Cornell, 1949), p. 16 ff.
 - 16 One of the best criticisms of the "appearance" (or "sense-datum") terminology was O. K. Bouwsma's "Moore's Theory of Sense-Data," in *The Philosophy of G. E. Moore*, ed. Schilpp, pp. 201–21. In *Perceiving: A Philosophical Study* (Cornell, 1957), I tried to call attention to certain facts about appearing which, I believe, Bouwsma may have overlooked.
 - 17 Augustine, *Contra academicos*, xi, 26; translated by Sister Mary Patricia Garvey as *Saint Augustine Against the Academicians* (Marquette, 1942); the quotations are from pp. 68–9.
 - 18 *Experience and Prediction*, p. 176.
 - 19 It may follow, however, that "the vaunted incorrigibility of the sense-datum language can be achieved only at the cost of its perfect utility as a means of communication" (Max Black, *Problems of Analysis* (Cornell, 1954), p. 66), and doubtless, as Black added, it would be "misleading, to say the least" to speak of a "language that cannot be communicated" – cf. Wilfrid Sellars, "Empiricism and the Philosophy of Mind" – but these points do affect the epistemological question at issue.
 - 20 This doctrine was modified in Lewis's later *An Analysis of Knowledge and Valuation* (Open Court, 1946) in a way which enabled him to preserve the theory of the given.
 - 21 *The Nature of Thought*, vol. II, pp. 269–70. Blanchard added, however, that "for all the ordinary purposes of life" we can justify some beliefs by showing that they cohere "with the system of present knowledge"; and therefore, he said, his theory should not be described as being "simply sceptical" (vol. II, p. 271). Cf. W. H. Werkmeister, *The Basis and Structure of Knowledge* (Harper, 1948), part II.
 - 22 Quoted by A. E. Murphy in "Dewey's Epistemology and Metaphysics," in P. A. Schilpp (ed.), *The Philosophy of John Dewey*, (Northwestern, 1939), p. 203. Dewey's theory of inquiry, however, was not intended to be an epistemology and he did not directly address himself to the questions with which we are here concerned.
 - 23 Cf. W. V. Quine, *Mathematical Logic* (Norton, 1940; rev. edn, Harvard, 1951), sec. 4.
 - 24 Ledger Wood, *The Analysis of Knowledge* (Princeton, 1941), p. 81; C. I. Lewis, *An Analysis of Knowledge and Valuation*, p. 334.
 - 25 Important steps toward solving them were taken by Nelson Goodman in "Sense and Certainty," *Philosophical Review* LXI (1952), pp. 160–7, and by Israel Scheffler in "On Justification and Commitment," *Journal of Philosophy*, LI (1954), pp. 180–90. The former paper is reprinted in *Philosophy of Knowledge*, ed. Roland Houde and J. P. Mullally (Lippincott, 1960), pp. 97–103.
 - 26 Alternatives to the general metaphor of the edifice are proposed by W. V. Quine in the introduction to *Methods of Logic* (Holt, 1950; rev. edn, 1959), in *From a Logical Point of View* (Harvard, 1953), and in *Word and Object* (Wiley, 1960).
 - 27 John Dewey, *Experience and Nature*, 2nd edn (Norton, 1929), p. 327.

Does Empirical Knowledge Have a Foundation?

Wilfrid Sellars

I have arrived at a stage in my argument which is, at least *prima facie*, out of step with the basic presuppositions of logical atomism. Thus, as long as *looking green* is taken to be the notion to which *being green* is reducible, it could be claimed with considerable plausibility that fundamental concepts pertaining to observable fact have that logical independence of one another which is characteristic of the empiricist tradition. Indeed, at first sight the situation is *quite* disquieting, for if the ability to recognize that *x* looks green presupposes the concept of *being green*, and if this in turn involves knowing in what circumstances to view an object to ascertain its color, then, since one can scarcely determine what the circumstances are without noticing that certain objects have certain perceptible characteristics – including colors – it would seem that one couldn't form the concept of *being green*, and, by parity of reasoning, of the other colors, unless he already had them.

Now, it just won't do to reply that to have the concept of green, to know what it is for something to be green, it is sufficient to respond, when one is *in point of fact* in standard conditions, to green objects with the vocable "This is green." Not only must the conditions be of a sort that is appropriate for determining the color of an object by looking, the subject must *know* that conditions of this sort *are* appropriate. And while this does not imply that one must have concepts before one has them, it does imply that one can have the concept

of green only by having a whole battery of concepts of which it is one element. It implies that while the process of acquiring the concept green may indeed do – involve a long history of acquiring *piecemeal* habits of response to various objects in various circumstances, there is an important sense in which one has *no* concept pertaining to the observable properties of physical objects in Space and Time unless one has them all – and, indeed, as we shall see, a great deal more besides.

...

One of the forms taken by the Myth of the Given is the idea that there is, indeed *must be*, a structure of particular matter of fact such that (a) each fact can not only be noninferentially known to be the case, but presupposes no other knowledge either of particular matter of fact, or of general truths; and (b) such that the noninferential knowledge of facts belonging to this structure constitutes the ultimate court of appeals for all factual claims – particular and general – about the world. It is important to note that I characterized the knowledge of fact belonging to this stratum as not only noninferential, but as presupposing no knowledge of other matter of fact, whether particular or general. It might be thought that this is a redundancy, that knowledge (not belief or conviction, but knowledge) which logically presupposes knowledge of other facts *must be* inferential. This, however, as I hope to show, is itself an episode in the Myth.

Now, the idea of such a privileged stratum of fact is a familiar one, though not without its difficulties. Knowledge pertaining to this level is *non-inferential*, yet it is, after all, *knowledge*. It is *ultimate*, yet it has *authority*. The attempt to

Originally published in H. Feigl and M. Scriven (eds). *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minnesota Studies In the Philosophy of Science, vol. 1 (Minneapolis: University of Minnesota Press, 1956), pp. 293–300.

make a consistent picture of these two requirements has traditionally taken the following form:

Statements pertaining to this level, in order to “express knowledge” must not only be made, but, so to speak, must be worthy of being made, *credible*, that is, in the sense of worthy of credence. Furthermore, and this is a crucial point, they must be made in a way which *involves* this credibility. For where there is no connection between the making of a statement and its authority, the assertion may express *conviction*, but it can scarcely be said to express knowledge.

The authority – the credibility – of statements pertaining to this level cannot exhaustively consist in the fact that they are supported by *other* statements, for in that case all *knowledge* pertaining to this level would have to be inferential, which not only contradicts the hypothesis, but flies in the face of good sense. The conclusion seems inevitable that if some statements pertaining to this level are to express *noninferential* knowledge, they must have a credibility which is not a matter of being supported by other statements. Now there does seem to be a class of statements which fill at least part of this bill, namely such statements as would be said to *report observations*, thus, “This is red.” These statements, candidly made, have authority. Yet they are not expressions of inference. How, then, is this authority to be understood?

Clearly, the argument continues, it springs from the fact that they are made in just the circumstances in which they are made, as is indicated by the fact that they characteristically, though not necessarily or without exception, involve those so-called token-reflexive expressions which, in addition to the tenses of verbs, serve to connect the circumstances in which a statement is made with its sense. (At this point it will be helpful to begin putting the line of thought I am developing in terms of the *fact-stating* and *observation-reporting* roles of certain sentences.) Roughly, two verbal performances which are tokens of a non-token-reflexive sentence can occur in widely different circumstances and yet make the same statement; whereas two tokens of a token-reflexive sentence can make the same statement only if they are uttered in the same circumstances (according to a relevant criterion of sameness). And two tokens of a sentence, whether it contains a token-reflexive expression – over and above a

tensed verb – or not, can make the same *report* only if, made in all candor, they express the *presence* – in *some* sense of “presence” – of the state of affairs that is being reported; if, that is, they stand in that relation to the state of affairs, whatever the relation may be, by virtue of which they can be said to formulate observations of it.

It would appear, then, that there are two ways in which a sentence token can have credibility: (1) The authority may accrue to it, so to speak, from above, that is, as being a token of a sentence type *all* the tokens of which, in a certain use, have credibility, e.g. “ $2 + 2 = 4$.” In this case, let us say that token credibility is inherited from type authority. (2) The credibility may accrue to it from the fact that it came to exist in a certain way in a certain set of circumstances, e.g. “This is red.” Here token credibility is not derived from type credibility.

Now, the credibility of *some* sentence types appears to be *intrinsic* – at least in the limited sense that it is *not* derived from other sentences, type or token. This is, or seems to be, the case with certain sentences used to make analytic statements. The credibility of *some* sentence types accrues to them by virtue of their logical relations to other sentence types, thus by virtue of the fact that they are logical consequences of more basic sentences. It would seem obvious, however, that the credibility of empirical sentence types cannot be traced without remainder to the credibility of other sentence types. And since no empirical sentence type appears to have *intrinsic* credibility, this means that credibility must accrue to *some* empirical sentence types by virtue of their logical relations to certain sentence tokens, and, indeed, to sentence tokens the authority of which is not derived, in its turn, from the authority of sentence types.

The picture we get is that of their being two *ultimate* modes of credibility: (1) The intrinsic credibility of analytic sentences, which accrues to tokens as being tokens of such a type; (2) the credibility of such tokens as “express observations,” a credibility which flows from tokens to types.

Let us explore this picture, which is common to all traditional empiricisms, a bit further. How is the authority of such sentence tokens as “express observational knowledge” to be understood? It has been tempting to suppose that in spite of the obvious differences which exist between “observa-

tion reports” and “analytic statements,” there is an essential similarity between the ways in which they come by their authority. Thus, it has been claimed, not without plausibility, that whereas *ordinary* empirical statements can be *correctly* made without being *true*, observation reports resemble analytic statements in that being correctly made is a sufficient as well as necessary condition of their truth. And it has been inferred from this – somewhat hastily, I believe – that “correctly making” the report “This is green” is a matter of “following the rules for the use of ‘this,’ ‘is’ and ‘green.’”

Three comments are immediately necessary:

(1) First a brief remark about the term “report.” In ordinary usage a report is a report made *by* someone *to* someone. To make a report is *to do* something. In the literature of epistemology, however, the word “report” or “*Konstatierung*” has acquired a technical use in which a sentence token can play a reporting role (a) without being an *overt* verbal performance, and (b) without having the character of being “by someone to someone” – even oneself. There is, of course, such a thing as “talking to oneself” – *in foro interno* – but, as I shall be emphasizing in the closing stages of my argument, it is important not to suppose that all “covert” verbal episodes are of this kind.

(2) My second comment is that while *we* shall not assume that because “reports” *in the ordinary sense* are *actions*, “reports” in the sense of *Konstatierungen* are also actions, the line of thought we are considering treats them as such. In other words, it interprets the correctness of *Konstatierungen* as analogous to the rightness of actions. Let me emphasize, however, that not all *ought* is *ought to do*, nor all correctness the correctness of actions.

(3) My third comment is that if the expression “following a rule” is taken seriously, and is not weakened beyond all recognition into the bare notion of exhibiting a uniformity – in which case the lightning, thunder sequence would “follow a rule” – then it is the knowledge or belief that the circumstances are of a certain kind, and not the mere fact that they *are* of this kind, which contributes to bringing about the action.

In the light of these remarks it is clear that *if* observation reports are construed as *actions*, *if* their correctness is interpreted as the correctness of an *action*, and *if* the authority of an observation report is construed as the fact that making it is “following a rule” in the proper sense of this phrase, *then* we are face to face with givenness in its most straight-

forward form. For these stipulations commit one to the idea that the authority of *Konstatierungen* rests on nonverbal episodes of awareness – awareness *that* something is the case, e.g. *that this is green* – which nonverbal episodes have an intrinsic authority (they are, so to speak, “self-authenticating”) which the *verbal* performances (the *Konstatierungen*) properly performed “express.” One is committed to a stratum of authoritative nonverbal episodes (“awareness”) the authority of which accrues to a superstructure of *verbal actions*, provided that the expressions occurring in these actions are properly *used*. These self-authenticating episodes would constitute the tortoise on which stands the elephant on which rests the edifice of empirical knowledge. The essence of the view is the same whether these intrinsically authoritative episodes are such items as the awareness that a certain sense content is green or such items as the awareness that a certain physical object looks to someone to be green.

But what is the alternative? We might begin by trying something like the following: An overt or covert token of “This is green” in the presence of a green item is a *Konstatierung* and expresses observational knowledge if and only if it is a manifestation of a tendency to produce overt or covert tokens of “This is green” – given a certain set – if and only if a green object is being looked at in standard conditions. Clearly on this interpretation the occurrence of such tokens of “This is green” would be “following a rule” only in the sense that they are instances of a uniformity, a uniformity differing from the lightning-thunder case in that it is an acquired causal characteristic of the language user. Clearly the above suggestion, which corresponds to the “thermometer view” criticized by Professor Price, and which we have already rejected elsewhere, won’t do as it stands. Let us see, however, if it can’t be revised to fit the criteria I have been using for “expressing observational knowledge.”

The first hurdle to be jumped concerns the *authority* which, as I have emphasized, a sentence token must have in order that it may be said to express knowledge. Clearly, on this account the only thing that can remotely be supposed to constitute such authority is the fact that one can infer the presence of a green object from the fact that someone makes this report. As we have already noticed, the correctness of a report does not have to be construed as the rightness of an *action*. A report can be correct as being an instance of a

general mode of behavior which, in a given linguistic community, it is reasonable to sanction and support.

The second hurdle is, however, the decisive one. For we have seen that to be the expression of knowledge, a report must not only *have* authority, this authority must *in some sense* be recognized by the person whose report it is. And this is a steep hurdle indeed. For if the authority of the report "This is green" lies in the fact that the existence of green items appropriately related to the perceiver can be inferred from the occurrence of such reports, it follows that only a person who is able to draw this inference, and therefore who has not only the concept *green*, but also the concept of uttering "This is green" – indeed, the concept of certain conditions of perception, those which would correctly be called "standard conditions" – could be in a position to token "This is green" in recognition of its authority. In other words, for a *Konstatierung* "This is green" to "express observational knowledge," not only must it be a *symptom* or *sign* of the presence of a green object in standard conditions, but the perceiver must know that tokens of "This is green" *are* symptoms of the presence of green objects in conditions which are standard for visual perception.

Now it might be thought that there is something obviously absurd in the idea that before a token uttered by, say, Jones could be the expression of observational knowledge, Jones would have to know that overt verbal episodes of this kind are reliable indicators of the existence, suitably related to the speaker, of green objects. I do not think that it is. Indeed, I think that something very like it is true. The point I wish to make now, however, is that if it *is* true, then it follows, as a matter of simple logic, that one couldn't have observational knowledge of *any* fact unless one knew many *other* things as well. And let me emphasize that the point is not taken care of by distinguishing between *knowing how* and *knowing that*, and admitting that observational knowledge requires a lot of "know how." For the point is specifically that observational knowledge of any particular fact, e.g. that this is green, presupposes that one knows general facts of the form *X is a reliable symptom of Y*. And to admit this requires an abandonment of the traditional empiricist idea that observational knowledge "stands on its own feet." Indeed, the suggestion would be anathema to traditional empiricists for the obvious reason that by making observational knowledge *presuppose* knowledge of

general facts of the form *X is a reliable symptom of Y*, it runs counter to the idea that we come to know general facts of this form only *after* we have come to know by observation a number of particular facts which support the hypothesis that X is a symptom of Y.

And it might be thought that there is an obvious regress in the view we are examining. Does it not tell us that observational knowledge at time *t* presupposes knowledge of the form *X is a reliable symptom of Y*, which presupposes *prior* observational knowledge, which presupposes *other* knowledge of the form *X is a reliable symptom of Y*, which presupposes still other, and *prior*, observational knowledge, and so on? This charge, however, rests on too simple, indeed a radically mistaken, conception of what one is saying of Jones when one says that he *knows* that *p*. It is not just that the objection supposes that knowing is an *episode*; for clearly there are episodes which we can correctly characterize as knowings, in particular, *observings*. The essential point is that in characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says.

Thus, all that the view I am defending requires is that no tokening by *S* *now* of "This is green" is to count as "expressing observational knowledge" unless it is also correct to say of *S* that he *now* knows the appropriate fact of the form *X is a reliable symptom of Y*, namely that (and again I oversimplify) utterances of "This is green" are reliable indicators of the presence of green objects in standard conditions of perception. And while the correctness of this statement about Jones requires that Jones could *now* cite prior particular facts as evidence for the idea that these utterances *are* reliable indicators, it requires only that it is correct to say that Jones *now* knows, thus remembers, that these particular facts *did* obtain. It does not require that it be correct to say that at the time these facts did obtain he *then knew* them to obtain. And the regress disappears.

Thus, while Jones' ability to give inductive reasons *today* is built on a long history of acquiring and manifesting verbal habits in perceptual situations, and, in particular, the occurrence of verbal episodes, e.g. "This is green," which is superficially like those which are later properly said to express observational knowledge, it does not require that any episode in this prior time be

characterizeable as expressing knowledge. (At this point, the reader should reread the opening section of this chapter.)

The idea that observation “strictly and properly so-called” is constituted by certain self-authenticating nonverbal episodes, the authority of which is transmitted to verbal and quasi-verbal performances when these performances are made “in conformity with the semantical rules of the language,” is, of course, the heart of the Myth of the Given. For the *given*, in epistemological tradition, is what is *taken* by these self-authenticating episodes. These “takings” are, so to speak, the unmoved movers of empirical knowledge, the “knowings in presence” which are presupposed by all other knowledge, both the knowledge of general truths and the knowledge “in absence” of other particular matters of fact. Such is the framework in which traditional empiricism makes its characteristic claim that the perceptually given is the foundation of empirical knowledge.

If I reject the framework of traditional empiricism, it is not because I want to say that empirical knowledge has *no* foundation. For to put it this

way is to suggest that it is really “empirical knowledge so-called,” and to put it in a box with rumors and hoaxes. There is clearly *some* point to the picture of human knowledge as resting on a level of propositions – observation reports – which do not rest on other propositions in the same way as other propositions rest on them. On the other hand, I do wish to insist that the metaphor of “foundation” is misleading in that it keeps us from seeing that if there is a logical dimension in which other empirical propositions rest on observation reports, there is another logical dimension in which the latter rest on the former.

Above all, the picture is misleading because of its static character. One seems forced to choose between the picture of an elephant which rests on a tortoise (What supports the tortoise?) and the picture of a great Hegelian serpent of knowledge with its tail in its mouth (Where does it begin?). Neither will do. For empirical knowledge, like its sophisticated extension, science, is rational, not because it has a *foundation* but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once.

Epistemic Principles

Wilfrid Sellars

I

The explication of knowledge as 'justified true belief', though it involves many pitfalls to which attention has been called in recent years, remains the orthodox or classical account and is, I believe, essentially sound. Thus, in the present lecture I shall assume that it can be formulated in such a way as to be immune from the type of counter-examples with which it has been bombarded since Gettier's pioneering paper in *Analysis* and turn my attention to another problem which has dogged its footsteps since the very beginning. This problem can be put in the form of two questions. If knowledge is justified true belief, how can there be such a thing as self-evident knowledge? And if there is no such thing as self-evident knowledge, how can any true belief be, in the relevant sense, justified?

But first let us beat about in the neighboring fields, perhaps to scare up some game, but, in any case, to refamiliarize ourselves with the terrain. Thus, are there not occasions on which a person can be said to be justified in believing something which he would not appropriately be said to know? Presumably, to be justified in believing something is to have good reasons for believing it, as contrasted with its contradictory. But *how* good? Adequate? Conclusive? If adequate, adequate for what? If conclusive, the conclusion of what is at stake?

We are all familiar with Austin's point concerning the performative character of 'I know'. We are

also familiar with the fact that, whereas to say 'I promise to do *A*' is, other things being equal, to promise to do *A*, to say 'I know that-*p*' is not, other things being equal, to know that-*p*. Chisholm's distinction between the *strict* and the *extended* sense of 'performative utterance' is helpful in this connection. According to Chisholm,

An utterance beginning with 'I want' is not performative in [the] strict sense, for it cannot be said to be an 'act' of wanting. But 'I want' is often used to accomplish what one might accomplish by means of the strict performative 'I request'. Let us say, then, that 'I want' may be a 'performative utterance' in an *extended sense* of the latter expression.¹

He asks in which, if either, of these senses an utterance of 'I know' may be performative. After reminding us that 'I know' is not performative in the strict sense of the term, he allows that '[it] is often used to accomplish what one may accomplish by the strict performative "I guarantee" or "I give you my word"' and 'hence may be performative in an extended sense of the term'.²

He argues, however, that 'I know' is not always a substitute for 'I guarantee', pointing out that:

Just as an utterance of 'I want' may serve *both* to say something about me and to get you to do something, an utterance of 'I know' may serve both to say something about me and to provide you with guarantees. To suppose that the performance of the nondescriptive function is inconsistent with the simultaneous performance of the descriptive function might be called,

Originally published in H. Castaneda (ed.), *Action, Knowledge, and Reality* (Indianapolis: Bobbs-Merrill, 1975), pp. 332-49.

therefore, an example of the *performative fallacy*.³

I think that Chisholm is quite right about this. On the other hand, it seems to me that he overlooks the possibility of a connection between 'I know' and 'I guarantee' other than the one he considers. 'I know that-*p*' might be related to 'I guarantee that-*p*' not just as an autobiographical description which on occasion performs the same role as the latter but as one which contains a reference to guaranteeing in its very meaning. Is it not possible to construe 'I know that-*p*' as essentially equivalent to '*p*, and I have reasons good enough to support a guarantee' (i.e., to say 'I guarantee' or 'You can rely on my statement')? Such an account would enable us to recognize a performative element in the very meaning of the verb 'to know' without construing 'I know' as a performative in the strict sense. It would also preserve the symmetry between first person and other person uses of the verb 'to know' which seems to be a pre-analytic datum. Thus, 'He knows that-*p*' would entail 'He has reasons good enough to support a guarantee that-*p*'.⁴

Furthermore, this account would enable us to appreciate the *context dependence* of the adequacy involved. Reasons which might be adequately good to justify a guarantee on one occasion might not be adequate to justify a guarantee on another. Again, the presence of such a performative element in the very meaning of the verb 'to know' would account for the fact (if it is a fact) that we rarely think in terms of 'I know' in purely self-directed thinkings; that we rarely have thoughts of the form 'I know that-*p*' unless the question of a possible guarantee to someone other than ourselves has arisen. Of course, we *can* 'tell ourselves' that we know something, but, then, so can we be said to make promises to ourselves.

II

Yet even after justice has been done, perhaps along the above lines, to the performative element in the meaning of the verb 'to know', it seems to me that we must recognize a closely related use of this expression which, though it may have implications concerning action, is not in any of the above senses performative. For once the *ethical* issue of how good one's reasons for a belief must be in order to justify *giving a guarantee* is solved, there remains

the problem of how good reasons must be to justify believing that-*p*, where to believe that-*p* is obviously not an *action*, let alone a performatory action in either the strict or the extended sense.

Confronted by this question, we are tempted to set apart a class of cases in which the reasons are not only good enough to justify believing that-*p* but good enough to make it absurd *not* to believe that-*p* (or, perhaps, to believe its contradictory). It is perhaps, some such concept as this which is (in addition to the truth condition) the non-performative core of the meaning of the verb 'to know'.

I think the above discussion has served its primary purpose by highlighting the concept of having good reasons for believing that-*p*. For the solution of the problem which was posed in my opening remarks hinges ultimately on a distinction between two ways in which there can be, and one can have, good reasons for believing that-*p*.⁵

Now one pattern for justifying a belief in terms of good reasons can be called *inferential*. Consider the schema:

p;
So, I have good reasons, all things considered,
for believing *q*.

On reflection, this schema tends to expand into:

I have good reasons, all things considered, for believing *p*;
So, *p*;
So, I have good reasons, all things considered,
for believing *q*.

Further reflection suggests that arguments conforming to this schema have a suppressed premise. What might it be? Consider the following expanded schema:

I have, all things considered, good reasons for believing *p*;
So, *p*;
p logically implies *q*;
So, I have, all things considered, good reasons for believing *q*.

The line of thought thus schematically represented would seem to involve the principle,

Logical implication transmits reasonableness.

In cases of this type, we are tempted to say, we have *derivative* good reasons, all things considered, for believing q . We say, in other words, that the reasonableness of believing q is 'inferential'.

Notice that the above line of thought is obviously an oversimplification, undoubtedly in several respects. In particular, it is important to note that if I have independent grounds for believing $not-q$, I may decide that I do *not* have good reasons, all things considered, for believing that p . After all, if p implies q , $not-q$ equally implies $not-p$. Yet in spite of its oversimplifications, the above train of thought takes us nearer to the distinctions necessary to solve our problem.

I have been considering the case where one proposition, p , logically implies another, q , and have claimed, with the above qualifications, that logical implication transmits reasonableness. Perhaps we can also take into account, with trepidation, 'probabilistic' implication, which would give us the following schema:

It is reasonable, all things considered, to believe
 p ;
 So, p ;
 p probabilistically implies q to a high degree;
 So, all things considered, it is reasonable to believe q .

Probabilistic justification of beliefs in accordance with this pattern would, presumably, be illustrated by inductive arguments and theoretical explanations. In each case, we move from a premise of the form:

It is reasonable, all things considered, to believe
 E ,

where ' E ' formulates the evidence, to a conclusion of the form:

It is reasonable, all things considered, to believe
 H ,

where ' H ' formulates in the first case a law-like statement and in the second case a body of theoretical assumptions.

III

As has been pointed out since time immemorial, it is most implausible to suppose that all epistemic

justification is inferential, at least in the sense of conforming to the patterns described above. Surely, it has been argued, there must be beliefs which we are justified in holding on grounds other than that they can be correctly inferred, inductively or deductively, from other beliefs which we are justified in holding. In traditional terms, if there is to be *inferential* knowledge, must there not be *non-inferential* knowledge – beliefs, that is, the reasonableness of which does not rest on the reasonableness of beliefs which logically or probabilistically imply them?

We are clearly in the neighborhood of what has been called the 'self-evident', the 'self-certifying', in short, of 'intuitive knowledge'. It is in this neighborhood that we find what has come to be called the *foundational* picture of human knowledge. According to this picture, beliefs which have inferential reasonableness ultimately rely for their authority on a stratum of beliefs which are, in some sense, self-certifying. The reasonableness of moves from the level of the self-evident to higher levels would involve the principles of logic (deductive and inductive) and, perhaps, certain additional principles which are *sui generis*. They would have in common the character of transmitting authoritativeness from lower-level beliefs to higher-level beliefs.

IV

Let us reflect on the concept of such a foundational level of knowledge. It involves the concept of beliefs which are *reasonable*, which have *epistemic authority* or *correctness*, but which are not reasonable or authoritative by virtue of the fact that they are beliefs in propositions which are implied by other propositions which it is reasonable to believe. Let us label them, for the moment, 'non-inferentially reasonable beliefs'.

How can there be such beliefs? For the concept of a *reason* seems so clearly tied to that of an *inference* or *argument* that the concept of non-inferential reasonableness seems to be a *contradictio in adjecto*. Surely, we are inclined to say, for a belief (or believing) to be reasonable, there must be a reason for the belief (or believing). And must not this reason be something other than the belief or believing for which it is the reason? And surely, we are inclined to say, to believe something *because* it is reasonable (to believe it) involves not only that there *be* a reason but that, in a relevant sense, one

has or is in possession of the reason. Notice that I have deliberately formulated these expostulations in such a way as to highlight the ambiguities involved when one speaks of reasonable beliefs.

In attempting to cope with these challenges, I shall leave aside problems pertaining to inferential and non-inferential reasonableness in logic and mathematics and concentrate on the apparent need for 'self evidence' in the sphere of empirical matters of fact.

How might a self-justifying belief be construed? One suggestion, modified from Chisholm's *Theory of Knowledge*,⁶ is to the effect that the justification of such beliefs has the form,

What justifies me in claiming that my belief that *a* is *F* is reasonable is simply the fact that *a* is *F*.

But this seems to point to the existence of inferences of the form,

It is a fact that *a* is *F*;
So, it is reasonable to believe that *a* is *F*,

and one might begin to wonder what principle authorizes this inference.

Something, clearly, has gone wrong. In order for any such argument to do the job, its premise would have to have authority; it would have to be something which it is reasonable to believe. But if we modify the schema to take this into account, it becomes:

It is reasonable to believe it to be a fact that *a* is *F*;
So, it is reasonable to believe that *a* is *F*,

which, in virtue of the equivalence of

believing *a* to be *F*

with

believing it to be a fact that *a* is *F*,

is obviously unilluminating.

V

Now many philosophers who have endorsed a concept of intuitive knowledge are clearly committed to the position that there is a level of

cognition more basic than *believing*. This more basic level would consist of a sub-conceptual⁷ awareness of certain facts. In terms of the framework that I have sketched elsewhere, there would be a level of cognition more basic than *thinkings* or tokenings of sentences in Mentalese – more basic, in fact, than symbolic activity, literal or analogical. It would be a level of cognition unmediated by concepts; indeed it would be the very *source* of concepts in some such way as described by traditional theories of abstraction. It would be 'direct apprehension' of facts; their 'direct presence' to the mind.⁸

Schematically we would have,

It is a fact (which I directly apprehend) that *a* is *F*;
So, it is reasonable to have the *conceptual belief* that *a* is *F*.

This multiplication of distinctions raises two serious problems: (1) What sort of entities are *facts*? Do they belong to the real (extra-conceptual) order? That 'fact' is roughly a synonym for 'truth', and 'true' is appropriately predicated of conceptual items (in overt speech or Mentalese) should give pause for thought.

Then there is also the question: (2) How is 'direct apprehension' to be understood? If the *apprehending* is distinguishable from the *apprehended*, is it not also 'separable'? Might not *apprehending* occur without any *fact* being *apprehended*? If so, an 'apprehending that-*p*' might not be an apprehending of the fact that-*p*. Hitting, in baseball, implies that something is hit. 'Swinging' does not. To *hit* is to *swing successfully*. Of course, 'apprehend', like 'see', is, in its ordinary sense, an achievement word. But does this not mean that, as in the case of 'see', there is a place for 'ostensibly apprehending', i.e., *seeming to apprehend*, a concept which does not imply achievement?

Many who use the metaphor 'to see' in intellectual contexts overlook the fact that in its literal sense 'seeing' is a term for a *successful* conceptual activity which contrasts with 'seeming to see'. No piling on of additional metaphors (e.g., 'grasping', which implies an object grasped) can blunt this fact. Now the distinction between *seeing* and merely *seeming to see* implies a criterion. To rely on the metaphors of 'apprehending' or 'presence of the object' is to obscure the need of criteria for distinguishing between 'knowing' and 'seeming to know', which ultimately define what it means to

speak of knowledge as a *correct* or well-founded *thinking* that something is the case.

If so, to know that we have apprehended a fact, we would have to know that the criteria which distinguish *apprehending* from *seeming to apprehend* were satisfied. In short, I suspect that the notion of a non-conceptual 'direct apprehension' of a 'fact' provides a merely verbal solution to our problem. The regress is stopped by an *ad hoc* regress-stopper. Indeed, the very metaphors which promised the sought-for foundation contain within themselves a dialectical moment which takes us beyond them.

VI

What is the alternative? I suggest that the key to our problem is provided by the Verbal Behaviorist model, developed elsewhere. It is, we have seen, a simple, indeed radically over-simplified, model, but it will provide us, I believe, with the outline of a strategy for getting out of the classical labyrinth.

According to this model, it will be remembered, the primary sense of

The thought occurred to Jones that snow is white

is

Jones said 'snow is white',

where the verb 'to say' was stripped of some of its ordinary implications and roughly equated with 'to utter words candidly as one who knows the language'. In particular, it was purged of the illocutionary and perlocutionary forces which Austin and Grice find so central to their theory of meaning. 'To say', in this sense, was also equated with 'thinking-out-loud'.

According to the *VB*, as I describe him, we must also introduce, in order to take account of those cases where one thinks silently, a *secondary* sense of

The thought occurred to Jones that snow is white,

in which it refers to a short-term proximate propensity to think-out-loud that snow is white. When this propensity is 'uninhibited', *one thinks-out-loud*, i.e., thinks in the primary sense of this

term (as construed by *VB*). There can be many reasons why, on a particular occasion, this propensity is inhibited. But, for our purposes, the most important is the general inhibition acquired in childhood when, after being taught to think-out-loud, one is trained not to be a 'babblor'. One might use the model of an on-off switch which gets into the wiring diagram when the child learns to keep his thoughts to himself.

I have argued elsewhere that yet another concept of 'having the thought occur to one that-*p*' can be introduced which stands to the second as the theoretical concept of electronic processes stands to the acquisition (and loss) of the power to attract iron filings (or a bell clapper) by a piece of soft iron in a coil of wire attached to an electric circuit. I argued that the classical concept of thought-episodes can be construed as part of a theoretical framework designed to explain the acquisition and loss of verbal propensities to think-out-loud. In approaching the problem of the status of non-inferential knowledge, however, I shall return to the *VB* model and concentrate, indeed, on the primary sense of having the thought occur to one that-*p*, i.e., think-out-loud that-*p*.

I have argued elsewhere that perceptual experience involves a sensory element which is in no way a form of thinking, however intimately it may be connected with thinking. This element consists of what I have variously called 'sense impressions', 'sensations', or 'sensa'. I argued that these items, properly construed, belong in a theoretical framework designed to explain:

- (a) the difference between merely thinking of (believing in the existence of) a perceptible state of affairs and *seeing* (or *seeming to see*) that such a state of affairs exists;
- (b) how it can seem to a person that there is a pink ice cube in front of him when there isn't one – either because there is something there which is either not pink or not cubical, or because there is nothing there and he is having a realistic hallucination.

I've explored problems pertaining to the nature and status of this sensory element on many occasions,⁹ but further exploration of this theme would leave no time for the problem at hand.

What is important for our purposes is that perceptual experience also involves a conceptual or propositional component – a 'thinking' in a suitably broad sense of this accordion term. In perception, the thought is caused to occur to one that, for

example, there is a pink ice cube in front of one. It is misleading to call such a thought a 'perceptual judgment' – for this implies question-answering activity of estimating, for example, the size of an object. (I judge that the room is ten feet tall.) Perhaps the best term is 'taking something to be the case'. Thus, on the occasion of sensing a certain color configuration, one takes there to be an object or situation of a certain description in one's physical environment.

Let us consider the case where

Jones sees there to be a red apple in front of him.

Given that Jones has learned how to use the relevant words in perceptual situations, he is justified in reasoning as follows:

I just thought-out-loud 'Lo! Here is a red apple' (no countervailing conditions obtain);
So, there is good reason to believe that there is a red apple in front of me.

Of course, the conclusion of this reasoning is not the *thinking* involved in his original perceptual experience. Like all justification arguments, it is a higher-order thinking. He did not originally *infer* that there is a red apple in front of him. Now, however, he is inferring from the character and context of his experience that it is veridical and that there is good reason to believe that there is indeed a red apple in front of him.

Notice that although the justification of the belief that there is a red apple in front of (Jones) is an inferential justification, it has the peculiar character that its essential premise asserts the occurrence of the very same belief in a specific context.¹⁰ It is this fact which gives the appearance that such beliefs are *self-justifying* and hence gives the justification the appearance of being *non-inferential*.

It is, as I see it, precisely this feature of the unique pattern of justification in question which, misinterpreted, leads Chisholm to formulate as his principle for the 'directly evident',

What justifies me in counting it as evident that *a* is *F* is simply the fact that *a* is *F*.¹¹

To be sure, Chisholm's examples of the 'directly evident' are not taken from the domain of *perceptual* beliefs, but rather, in true Cartesian spirit,

from one's knowledge about what is going on in one's mind at the present moment. Indeed, he rejects the idea that particular perceptual beliefs of the kind which I illustrated by my example of the red apple are ever directly evident.

On the other hand, though he does think that particular perceptual beliefs of this type can at best be *indirectly evident*, he does think that they can be *reasonable*. Should we say '*directly* reasonable'? I, of course, would answer in the affirmative. Yet it is not clear to me that Chisholm would be happy with this suggestion. If (as he should) he has at the back of his mind the reasoning;

There (visually) appears to me to be a red apple here;

So, it is reasonable for me (to believe) that there is a red apple here,

then he should not object to speaking of the reasonableness in question as 'direct', for the premise does not contain a predicate of epistemic evaluation. If, on the other hand (as he should not), he has at the back of his mind the following reasoning,

It is *evident* to me that there (visually) appears to me to be a red apple here;

So, it is *reasonable* for me (to believe) that there is a red apple here,

we could expect him to object to speaking of this reasonableness as 'direct'.

This tension sets the stage for a corresponding comment on Chisholm's third epistemic principle, which concerns the case where what we visually take to be the case is the presence of something having a 'sensible characteristic *F*' (where '*F*' ranges over the familiar Aristotelian list of proper and common sensibles). The principle reads as follows:

(C) If there is a certain sensible characteristic *F* such that *S* believes that he perceives something to be *F*, then it is *evident* to *S* that he is perceiving something to have that characteristic *F*, and also *evident* that there is something that is *F*.

I shall not pause to quibble over such matters as whether, in the light of Chisholm's definition of 'evident', it can ever be evident to me that I am perceiving something to be pink or that something

in front of me is pink – even if the claim is limited to the facing side. A high degree of reasonableness will do. The point which I wish to stress is that once again the question arises, does Chisholm think of the evidence involved in the principles as ‘direct’ or ‘indirect’? This time it is clear that he thinks of it as *indirect*. As I see it, then, he has at the back of his mind the following reasoning:

It is *evident* to me that there appears to me to be a pink object here;
So, it is *evident* to me that I perceive a pink object to be here and *evident* to me that there is a pink object here.

The contrasting reasoning would be:

There appears to me to be a pink object here;
So, it is *evident* to me that I perceive a pink object to be here and *evident* to me that there is a pink object here.

Now I suspect that what has misled Chisholm is the fact that if I were to argue,

There appears to me to be a pink cube here;
So, it is highly reasonable for me (to believe) that there is a pink object here,

a skeptic could be expected to challenge me by asking ‘What right have you to accept your conclusion, unless you have a right to accept the premise? Are you not implying that *you know* that there appears to you to be a pink object here; and must not this claim be a tacit *premise* in your argument?’ But, surely, the skeptic would just be mistaken – not, indeed, in asserting that in some sense I *imply* that I know that there appears to me to be a pink object here, but in asserting that this implication must be taken to be a premise in my reasoning, if it is to be *valid*, and, hence, if the corresponding epistemic principle is to be *true*. But in that case, the latter principle would be *not* Chisholm’s (C), but rather:

(C') *If it is evident to S that there is a certain sensible characteristic F...*

The larger import of the above reply to the skeptic will be sketched in my concluding remarks. For the moment, let me say that from my point of view something very like Chisholm’s principle (C) is sound but concerns the *direct* evidence (or, bet-

ter, *direct* high degree of reasonableness) of certain perceptual beliefs. Let me formulate it as follows:

(S) If there is a certain sensible characteristic *F* such that *S* believes that he perceives something to be *F*, then it is *evident* to *S* that there is something that is *F* and, hence, that he is perceiving something to be *F*.

Notice that I have reversed the relative position of the two clauses in the consequent as they appear in Chisholm’s principle. This is because, on my interpretation, the *core* of the principle is

(S1) If I ostensibly see there to be an *F* object here, then it is highly reasonable for me (to believe) that there is an *F* object here.

And the move to

(S2) If I ostensibly see there to be an *F* object here, then it is highly reasonable for me (to believe) *that I see* there to be an *F* object here

is justified by the conceptual tie between ‘ostensibly see’, ‘see’, and truth.

VII

Chisholm’s principle (C) and his other epistemic principles pertaining to perception and memory are themselves justified, as he sees it, by the fact that unless they, or something like them, are true, then there could be no such thing as perceptual knowledge to the effect, to use his example, that there is a cat on the roof. We have here a justification of the ‘this or nothing’ kind familiar to the Kantian tradition. The principles also seem, on occasion, to be treated as candidates for the status of synthetic *a priori* (and even, one suspects, self-evident) truth.

As I see it, on the other hand, these epistemic principles can be placed in a naturalistic setting and their authority construed in terms of the nature of concept formation and of the acquisition of relevant linguistic skills. The model which I have been using is, indeed, a very simple one, and I have largely limited my use of it to the epistemic authority of perceptual beliefs. But if the strategy which I have suggested is successful, it is a relatively simple matter to extend it to memory beliefs.

I have discussed the case of non-inferential knowledge of our own mental states in some detail, using this same general strategy, on a number of occasions.¹²

But, surely, it will be urged, facts about learning languages and acquiring linguistic skills are themselves empirical *facts*; and to know these facts involves perception, memory, indeed, all the epistemic activities the justification of which is at stake. Must we not conclude that any such account as I give of the principle that perceptual beliefs occurring in perceptual contexts are *likely to be true* is circular? It must, indeed, be granted that principles pertaining to the epistemic authority of perceptual and memory beliefs are not the sort of thing which *could* be arrived at by inductive reasoning from perceptual belief. But the best way to make this point is positive. *We have to be in this framework to be thinking and perceiving beings at all.* I suspect that it is this plain truth which is the real underpinning of the idea that the authority of epistemic principles rests on the fact that unless they were true we could not see that a cat is on the roof.

I pointed out a moment ago that we have to be in the framework of these (and other) principles to be thinking, perceiving, and, I now add, acting beings at all. But surely this makes it clear that the exploration of these principles is but part and parcel of the task of explicating the concept of a rational animal or, in *V/B* terms, of a language-using organism whose language is *about* the world in which it is *used*. It is only in the light of this larger task that the problem of the status of epistemic principles reveals its true meaning.

From the perspective of this larger task, the metaphor of 'foundation and superstructure' is

seen to be a false extrapolation, to use a Deweyan turn of phrase, from specific 'problematic situations' with respect to which it *is* appropriate. And when we concern ourselves, as Philosophy ultimately demands, with *how it is* with man and his world, as contrasted with the catch-as-catch-can procedures which generate man's awareness of himself and his world, surely we can say, as I wrote some fifteen years ago in an earlier essay on this topic,

There is clearly *some* point to the picture of human knowledge as resting on a level of propositions – observation reports – which do not rest on other propositions in the same way as other propositions rest on them. On the other hand, I do wish to insist that the metaphor of 'foundation' is misleading in that it keeps us from seeing that if there is a logical dimension in which other empirical propositions rest on observation reports, there is another logical dimension in which the latter rest on the former.

Above all, the picture is misleading because of its static character. One seems forced to choose between the picture of an elephant which rests on a tortoise (What supports the tortoise?) and the picture of a great Hegelian serpent of knowledge with its tail in its mouth (Where did it begin?). Neither will do. For empirical knowledge, like its sophisticated extension, science, is rational, not because it has a *foundation* but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once.¹³

Notes

1 R. M. Chisholm, *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1966), pp. 16–17.

2 *Ibid.*

3 *Ibid.*, p. 17.

4 Notice that the above account of the relation of 'I know' to a performative is not quite the same as Urmson's. According to the latter, as represented by Chisholm, to say that Mr Jones *knew* some proposition to be true is to say that Mr Jones was 'in a position in which he was entitled to say "I know"'. This account, as Chisholm points out, brings us back to the original problem of how the first person use of the verb is to be construed.

5 I have called attention elsewhere to the importance of distinguishing between questions concerning the reasonableness of believing that-*p* from questions concerning the reasonableness of 'acting on the proposition that-*p*', including guaranteeing that-*p*. The concept of acting on a proposition is clear only in simple cases, as when, for example, the proposition occurs as a premise in the agent's practical reasoning. When the agent takes probabilities into account, a far more complicated story is necessary to clarify the sense in which a person can be said to have acted on a given proposition. For a discussion of these problems, see my 'Induction as Vindication', *Philosophy of Science* 31 (1964), pp. 197–232.

- 6 Chisholm, *Theory of Knowledge*, p. 28. Chisholm's principle concerns 'what justifies us in counting it as evident that a is F '. But the 'evident' is defined on p. 22 as a special case of the 'reasonable'.
- 7 Where 'sub-conceptual' is far from being used as a pejorative term.
- 8 It is clearly some such position which is envisaged by many who explicitly reject the equation of knowledge with justified true belief. That it is *implicit* in Chisholm's position becomes clear not only when we reflect (as above) on what his principle concerning the directly evident might mean, but when we take into account his use of such phrases as 'state of affairs' that "presents itself to him" or that "is apprehended through itself" (Chisholm, *Theory of Knowledge*, p. 28) and his general commitment to a fact ontology (*ibid.*, chap. 7, *passim*), a 'fact', in the relevant sense, being a 'state of affairs which exists' (*ibid.*, p. 104). 'Exists' in this context should not be confused with the 'existential quantifier' but should be considered as a synonym for 'obtains'. It is obviously not self-contradictory to say that some states of affairs do not obtain.
- 9 Most recently in my *Science and Metaphysics* (London: Routledge and Kegan Paul, 1967), ch. 1, and in 'Science, Sense Impressions, and Sensa: A Reply to Cornman', *Review of Metaphysics* 25 (1971), which is a reply to Cornman's 'Sellars, Scientific Realism, and Sensa', *Review of Metaphysics* 24 (1970).
- 10 I called attention to this feature of the justification involved in 'non-inferential' knowledge in *Science, Perception and Reality* (London: Routledge and Kegan Paul, and New York: Humanities Press, 1963), chap. 3. Thus, I wrote '... one only knows what one has a right to think to be the case. Thus, to say that one directly knows that- p is to say that his right to the conviction that- p essentially involves the fact that the idea that- p occurred to the knower in a specific way' (*ibid.*, p. 88). I suggested that this 'kind of credibility' be called 'trans-level credibility', and the pattern of inference involved in the reasoning which mobilizes this credibility, 'trans-level inference'. A similar point was less clearly made in Sections 32-9 of my 'Empiricism and the Philosophy of Mind', in Herbert Feigl and Michael Scriven (eds), *Minnesota Studies in the Philosophy of Science* vol. 1 (Minneapolis: University of Minnesota Press, 1956). Reprinted as chapter 5 of my *Science, Perception and Reality*.
- 11 Chisholm, *Theory of Knowledge*, p. 28.
- 12 Most recently in my *Science and Metaphysics*, esp. pp. 71ff., 151ff.
- 13 'Empiricism and the Philosophy of Mind', sec. 38; quoted from *Science, Perception and Reality*, p. 170.

The Raft and the Pyramid

Ernest Sosa

Contemporary epistemology must choose between the solid security of the ancient foundationalist pyramid and the risky adventure of the new coherentist raft. Our main objective will be to understand, as deeply as we can, the nature of the controversy and the reasons for and against each of the two options. But first of all we take note of two underlying assumptions.

1 Two Assumptions

- (A1) Not everything believed is known, but nothing can be known without being at least believed (or accepted, presumed, taken for granted, or the like) in some broad sense. What additional requirements must a belief fill in order to be knowledge? There are surely at least the following two: (a) it must be true, and (b) it must be justified (or warranted, reasonable, correct, or the like).
- (A2) Let us assume, moreover, with respect to the second condition A1(b): first, that it involves a normative or evaluative property; and, second, that the relevant sort of justification is that which pertains to knowledge: epistemic (or theoretical) justification. Someone seriously ill may have two sorts of justification for believing he

will recover: the practical justification that derives from the contribution such belief will make to his recovery and the theoretical justification provided by the lab results, the doctor's diagnosis and prognosis, and so on. Only the latter is relevant to the question whether he knows.

2 Knowledge and Criteria

- a. There are two key questions of the theory of knowledge:
- (i) What do we know?
 - (ii) How do we know?
- The answer to the first would be a list of bits of knowledge or at least of types of knowledge, of the self, of the external world, of other minds, and so on. An answer to the second would give criteria (or canons, methods, principles, or the like) that would explain how we know whatever it is that we do know.
- b. In developing a theory of knowledge, we can begin either with a(i) or with a(ii). Particularism would have us begin with an answer to a(i) and only then take up a(ii) on the basis of that answer. Quite to the contrary, methodism would reverse that order. The particularist thus tends to be antisceptical on principle. But the methodist is as such equally receptive to scepticism and to the contrary. Hume, for example, was no less a methodist than Descartes. Each accepted, in effect, that only the obvious and what is proved deductively on its basis can possibly be known.

Originally published in *Midwest Studies in Philosophy*, Vol. 5: *Studies in Epistemology* (Minneapolis: University of Minnesota Press, 1980), pp. 3-25; an appendix to this paper is drawn from Ernest Sosa, "How Do You Know?" *American Philosophical Quarterly* 11 (1974), pp. 113-22.

- c. What, then, is the obvious? For Descartes it is what we know by intuition, what is clear and distinct, what is indubitable and credible with no fear of error. Thus for Descartes basic knowledge is always an infallible belief in an indubitable truth. All other knowledge must stand on that basis through deductive proof. Starting from such criteria (canons, methods, etc.), Descartes concluded that knowledge extended about as far as his contemporaries believed.¹ Starting from similar criteria, however, Hume concluded that both science and common sense made claims far beyond their rightful limits.
- d. Philosophical posterity has rejected Descartes's theory for one main reason: that it admits too easily as obvious what is nothing of the sort. Descartes's reasoning is beautifully simple: God exists; no omnipotent perfectly good being would descend to deceit; but if our common sense beliefs were radically false, that would represent deceit on His part. Therefore, our common sense beliefs must be true or at least cannot be radically false. But in order to buttress this line of reasoning and fill in details, Descartes appeals to various principles that appear something less than indubitable.
- e. For his part, Hume rejects all but a minuscule portion of our supposed common sense knowledge. He establishes first that there is no way to prove such supposed knowledge on the basis of what is obvious at any given moment through reason or experience. And he concludes, in keeping with this methodism, that in point of fact there really is no such knowledge.

3 Two Metaphors: The Raft and the Pyramid

Both metaphors concern the body or system of knowledge in a given mind. But the mind is of course a more complex marvel than is sometimes supposed. Here I do not allude to the depths plumbed by Freud, nor even to Chomsky's. Nor need we recall the labyrinths inhabited by statesmen and diplomats, nor the rich patterns of some novels or theories. We need look no further than the most common, everyday beliefs. Take, for instance, the belief that driving tonight will be dangerous. Brief reflection should reveal that any

of us with that belief will join to it several other closely related beliefs on which the given belief depends for its existence or (at least) its justification. Among such beliefs we could presumably find some or all of the following: that the road will be icy or snowy; that driving on ice or snow is dangerous; that it will rain or snow tonight; that the temperature will be below freezing; appropriate beliefs about the forecast and its reliability; and so on.

How must such beliefs be interrelated in order to help justify my belief about the danger of driving tonight? Here foundationalism and coherentism disagree, each offering its own metaphor. Let us have a closer look at this dispute, starting with foundationalism.

Both Descartes and Hume attribute to human knowledge an architectonic structure. There is a nonsymmetric relation of physical support such that any two floors of a building are tied by that relation: one of the two supports (or at least helps support) the other. And there is, moreover, a part with a special status: the foundation, which is supported by none of the floors while supporting them all.

With respect to a body of knowledge *K* (in someone's possession), foundationalism implies that *K* can be divided into parts *K*₁, *K*₂... such that there is some nonsymmetric relation *R* (analogous to the relation of physical support) which orders those parts in such a way that there is one – call it *F* – that bears *R* to every other part while none of them bears *R* in turn to *F*.

According to foundationalism, each piece of knowledge lies on a pyramid such as that shown in Figure 14.1. The nodes of such a pyramid (for a proposition *P* relative to a subject *S* and a time *t*) must obey the following requirements:

- a. The set of all nodes that succeed (directly) any given node must serve jointly as a base that properly supports that node (for *S* at *t*).
- b. Each node must be a proposition that *S* is justified in believing at *t*.
- c. If a node is not self-evident (for *S* at *t*), it must have successors (that serve jointly as a base that properly supports that node).

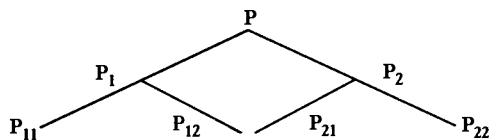


Figure 14.1

- d. Each branch of an epistemic pyramid must terminate.

For the foundationalist Descartes, for instance, each terminating node must be an indubitable proposition that S believes at *t* with no possibility of error. As for the nonterminal nodes, each of them represents inferential knowledge, derived by deduction from more basic beliefs.

Such radical foundationalism suffers from a fatal weakness that is twofold: (a) there are not so many perfectly obvious truths as Descartes thought; and (b) once we restrict ourselves to what is truly obvious in any given context, very little of one's supposed common sense knowledge can be proved on that basis. If we adhere to such radical foundationalism, therefore, we are just wrong in thinking we know so much.

Note that in citing such a "fatal weakness" of radical foundationalism, we favor particularism as against the methodism of Descartes and Hume. For we reject the methods or criteria of Descartes and Hume when we realize that they plunge us in a deep skepticism. If such criteria are incompatible with our enjoyment of the rich body of knowledge that we commonly take for granted, then as good particularists we hold on to the knowledge and reject the criteria.

If we reject radical foundationalism, however, what are we to put in its place? Here epistemology faces a dilemma that different epistemologists resolve differently. Some reject radical foundationalism but retain some more moderate form of foundationalism in favor of a radically different coherentism. Coherentism is associated with idealism – of both the German and the British variety – and has recently acquired new vigor and interest.

The coherentists reject the metaphor of the pyramid in favor of one that they owe to the positivist Neurath, according to whom our body of knowledge is a raft that floats free of any anchor or tie. Repairs must be made afloat, and though no part is untouchable, we must stand on some in order to replace or repair others. Not every part can go at once.

According to the new metaphor, what justifies a belief is not that it be an infallible belief with an indubitable object, nor that it have been proved deductively on such a basis, but that it cohere with a comprehensive system of beliefs.

4 A Coherentist Critique of Foundationalism

What reasons do coherentists offer for their total rejection of foundationalism? The argument that follows below summarizes much of what is alleged against foundationalism. But first we must distinguish between subjective states that incorporate a propositional attitude and those that do not. A propositional attitude is a mental state of someone with a proposition for its object: beliefs, hopes, and fears provide examples. By way of contrast, a headache does not incorporate any such attitude. One can of course be conscious of a headache, but the headache itself does not constitute or incorporate any attitude with a proposition for its object. With this distinction in the background, here is the antifoundationalist argument, which has two lemmas – a(iv) and b(iii) – and a principal conclusion.

- a. (i) If a mental state incorporates a propositional attitude, then it does not give us direct contact with reality, e.g., with pure experience, unfiltered by concepts or beliefs.
- (ii) If a mental state does not give us direct contact with reality, then it provides no guarantee against error.
- (iii) If a mental state provides no guarantee against error, then it cannot serve as a foundation for knowledge.
- (iv) Therefore, if a mental state incorporates a propositional attitude, then it cannot serve as a foundation for knowledge.
- b. (i) If a mental state does not incorporate a propositional attitude, then it is an enigma how such a state can provide support for any hypothesis, raising its credibility selectively by contrast with its alternatives. (If the mental state has no conceptual or propositional content, then what logical relation can it possibly bear to any hypothesis? Belief in a hypothesis would be a propositional attitude with the hypothesis itself as object. How can one depend logically for such a belief on an experience with no propositional content?)
- (ii) If a mental state has no propositional content and cannot provide logical support for any hypothesis, then it cannot serve as a foundation for knowledge.

- (iii) Therefore, if a mental state does not incorporate a propositional attitude, then it cannot serve as a foundation for knowledge.
- c. Every mental state either does or does not incorporate a propositional attitude.
- d. Therefore, no mental state can serve as a foundation for knowledge. (From a(iv), b(iii), and c.)

According to the coherentist critic, foundationalism is run through by this dilemma. Let us take a closer look.²

In the first place, what reason is there to think, in accordance with premise b(i), that only propositional attitudes can give support to their own kind? Consider practices – e.g., broad policies or customs. Could not some person or group be justified in a practice because of its consequences: that is, could not the consequences of a practice make it a good practice? But among the consequences of a practice may surely be found, for example, a more just distribution of goods and less suffering than there would be under its alternatives. And neither the more just distribution nor the lower degree of suffering is a propositional attitude. This provides an example in which propositional attitudes (the intentions that sustain the practice) are justified by consequences that are not propositional attitudes. That being so, is it not conceivable that the justification of belief that matters for knowledge be analogous to the objective justification by consequences that we find in ethics?

Is it not possible, for instance, that a belief that there is something red before one be justified in part because it has its origins in one's visual experience of red when one looks at an apple in daylight? If we accept such examples, they show us a source of justification that serves as such without incorporating a propositional attitude.

As for premise a(iii), it is already under suspicion from our earlier exploration of premise b(i). A mental state *M* can be nonpropositional and hence not a candidate for so much as truth, much less infallibility, while it serves, in spite of that, as a foundation of knowledge. Leaving that aside, let us suppose that the relevant mental state is indeed propositional. Must it then be infallible in order to serve as a foundation of justification and knowledge? That is so far from being obvious that it seems more likely false when compared with an analogue in ethics. With respect to beliefs, we may distinguish between their being true and their

being justified. Analogously, with respect to actions, we may distinguish between their being optimal (best of all alternatives, all things considered) and their being (subjectively) justified. In practical deliberation on alternatives for action, is it inconceivable that the most *eligible* alternative *not* be objectively the best, all things considered? Can there not be another alternative – perhaps a most repugnant one worth little if any consideration – that in point of fact would have a much better total set of consequences and would thus be better, all things considered? Take the physician attending to Frau Hitler at the birth of little Adolf. Is it not possible that if he had acted less morally, that would have proved better in the fullness of time? And if that is so in ethics, may not its likeness hold good in epistemology? Might there not be justified (reasonable, warranted) beliefs that are not even true, much less infallible? That seems to me not just a conceivable possibility, but indeed a familiar fact of everyday life, where observational beliefs too often prove illusory but no less reasonable for being false.

If the foregoing is on the right track, then the antifoundationalist is far astray. What has led him there?

As a diagnosis of the antifoundationalist argument before us, and more particularly of its second lemma, I would suggest that it rests on an Intellectualist Model of Justification.

According to such a model, the justification of belief (and psychological states generally) is parasitical on certain logical relations among propositions. For example, my belief (i) that the streets are wet, is justified by my pair of beliefs (ii) that it is raining, and (iii) that if it is raining, the streets are wet. Thus we have a structure such as this:

B(Q) is justified by the fact that B(Q) is grounded on (B(P), B(P \supset Q)).

And according to an Intellectualist Model, this is parasitical on the fact that

P and (P \supset Q) together logically imply Q.

Concerning this attack on foundationalism I will argue (a) that it is useless to the coherentist, since if the antifoundationalist dilemma impales the foundationalist, a form of it can be turned against the coherentist to the same effect; (b) that the dilemma would be lethal not only to foundationalism and coherentism but also to the very

possibility of substantive epistemology; and (c) that a form of it would have the same effect on normative ethics.

- a. According to coherentism, what justifies a belief is its membership in a coherent and comprehensive set of beliefs. But whereas being grounded on $B(P)$ and $B(P \subset Q)$ is a property of a belief $B(Q)$ that yields immediately the logical implication of Q and P and $(P \subset Q)$ as the logical source of that property's justificatory power, the property of being a member of a coherent set is not one that immediately yields any such implication.

It may be argued, nevertheless, (i) that the property of being a member of a coherent set would supervene in any actual instance on the property of being a member of a particular set a that is in fact coherent, and (ii) that this would enable us to preserve our Intellectualist Model, since (iii) the justification of the member belief $B(Q)$ by its membership in a would then be parasitical on the logical relations among the beliefs in a which constitute the coherence of that set of beliefs, and (iv) the justification of $B(Q)$ by the fact that it is part of a coherent set would then be *indirectly* parasitical on logical relations among propositions after all.

But if such an indirect form of parasitism is allowed, then the experience of pain may perhaps be said to justify belief in its existence parasitically on the fact that P logically implies $P!$ The Intellectualist Model seems either so trivial as to be dull, or else sharp enough to cut equally against both foundationalism and coherentism.

- b. If (i) only propositional attitudes can justify such propositional attitudes as belief, and if (ii) to do so they must in turn be justified by yet other propositional attitudes, it seems clear that (iii) there is no hope of constructing a complete epistemology, one which would give us, in theory, an account of what the justification of any justified belief would supervene on. For (i) and (ii) would rule out the possibility of a finite regress of justification.
- c. If only propositional attitudes can justify propositional attitudes, and if to do so they must in turn be justified by yet other propositional attitudes, it seems clear that there

is no hope of constructing a complete normative ethics, one which would give us, in theory, an account of what the justification of any possible justified action would supervene upon. For the justification of an action presumably depends on the intentions it embodies and the justification of these, and here we are already within the net of propositional attitudes from which, for the Intellectualist, there is no escape.

It seems fair to conclude that our coherentist takes his antifoundationalist zeal too far. His antifoundationalist argument helps expose some valuable insights but falls short of its malicious intent. The foundationalist emerges showing no serious damage. Indeed, he now demands equal time for a positive brief in defense of his position.

5 The Regress Argument

- a. The regress argument in epistemology concludes that we must countenance beliefs that are justified in the absence of justification by other beliefs. But it reaches that conclusion only by rejecting the possibility in principle of an infinite regress of justification. It thus opts for foundational beliefs justified in some noninferential way by ruling out a chain or pyramid of justification that has justifiers, and justifiers of justifiers, and so on *without end*. One may well find this too short a route to foundationalism, however, and demand more compelling reasons for thus rejecting an infinite regress as vicious. We shall find indeed that it is not easy to meet this demand.
- b. We have seen how even the most ordinary of everyday beliefs is the tip of an iceberg. A closer look below the surface reveals a complex structure that ramifies with no end in sight. Take again my belief that driving will be dangerous tonight, at the tip of an iceberg, (I), as presented in Figure 14.2. The immediate cause of my belief that driving will be hazardous tonight is the sound of raindrops on the windowpane. All but one or two members of the underlying iceberg are as far as they can be from my thoughts at the time. In what sense, then, do they form an iceberg whose tip breaks the calm surface of my consciousness?

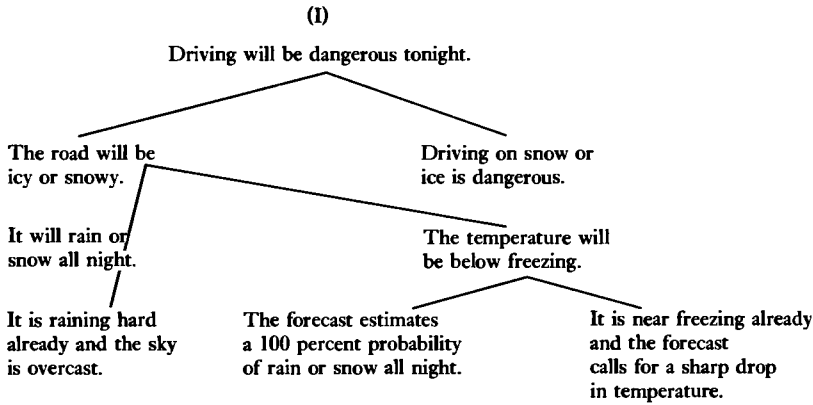


Figure 14.2

Here I will assume that the members of (I) are beliefs of the subject, even if unconscious or subconscious, that causally buttress and thus justify his prediction about the driving conditions.

Can the iceberg extend without end? It may appear obvious that it cannot do so, and one may jump to the conclusion that any piece of knowledge must be ultimately founded on beliefs that are *not* (inferentially) justified or warranted by other beliefs. This is a doctrine of *epistemic foundationalism*.

Let us focus not so much on the *giving* of justification as on the *having* of it. Can there be a belief that is justified in part by other beliefs, some of which are in turn justified by yet other beliefs, and so on without end? Can there be an endless regress of justification?

c. There are several familiar objections to such a regress:

(i) *Objection*: "It is incompatible with human limitations. No human subject could harbor the required infinity of beliefs."

Reply: It is mere presumption to fathom with such assurance the depths of the mind, and especially its unconscious and dispositional depths. Besides, our object here is the nature of epistemic justification in itself and not only that of such justification as is accessible to humans. Our question is not whether humans could harbor an infinite iceberg of justification. Our question is rather whether *any* mind, no matter how deep, could do so. Or is it ruled out *in principle* by the very nature of justification?

(ii) *Objection*: "An infinite regress is indeed ruled out in principle, for if justification were thus infinite how could it possibly end?"

Reply: (i) If the end mentioned is *temporal*, then why must there be such an end? In the first place, the subject may be eternal. Even if he is not eternal, moreover, why must belief acquisition and justification occur seriatim? What precludes an infinite body of beliefs acquired at a single stroke? Human limitations may rule this out for humans, but we have yet to be shown that it is precluded in principle, by the very nature of justification.

(ii) If the end mentioned is justificatory, on the other hand, then to ask how justification could possibly end is just to beg the question.

(iii) *Objection*: "Let us make two assumptions: first, that S's belief of *q* justifies his belief of *p* only if it works together with a justified belief on his part that *q* provides good evidence for *p*; and, second, that if S is to be justified in believing *p* on the basis of his belief of *q* and is to be justified in believing *q* on the basis of his belief of *r*, then S must be justified in believing that *r* provides good evidence for *p* via *q*. These assumptions imply that an actual regress of justification requires belief in an infinite proposition. Since no one (or at least no human) can believe an infinite proposition, no one (no human) can be a subject of such an actual regress."³

Reply: Neither of the two assumptions is beyond question, but even granting them both, it may still be doubted that the conclusion follows. It is true that each finitely complex belief of form “*r* provides good evidence for *p* via $q_1 \dots q_n$ ” will omit how some members of the full infinite regress are epistemically tied to belief of *p*. But that seems irrelevant given the fact that for each member *r* of the regress, such that *r* is tied epistemically to belief of *p*, there is a finite belief of the required sort (“*r* provides good evidence for *p* via $q_1 \dots q_n$ ”) that ties the two together. Consequently, there is no apparent reason to suppose – even granted the two assumptions – that an infinite regress will require a single belief in an infinite proposition, and not just an infinity of beliefs in increasingly complex finite propositions.

(iv) *Objection:* “But if it is allowed that justification extend infinitely, then it is too easy to justify any belief at all or too many beliefs altogether. Take, for instance, the belief that there are perfect numbers greater than 100. And suppose a mind powerful enough to believe every member of the following sequence:

- (σ1) There is at least one perfect number > 100
 - There are at least two perfect numbers > 100
 - There are at least three perfect numbers > 100

If such a believer has no other belief about perfect numbers save the belief that a perfect number is a whole number equal to the sum of its whole factors, then surely he is *not* justified in believing that there are perfect numbers greater than 100. He is quite unjustified in believing any of the members of sequence (σ1), in spite of the fact that a challenge to any can be met easily by appeal to its successor. Thus it cannot be allowed after all that justification extend infinitely, and an infinite regress is ruled out.”

Reply: We must distinguish between regresses of justification that are actual

and those that are merely potential. The difference is *not* simply that an actual regress is composed of actual beliefs. For even if all members of the regress are actual beliefs, the regress may still be *merely potential* in the following sense: while it is true that *if* any member *were* justified then its predecessors *would* be, still none is in fact justified. Anyone with our series of beliefs about perfect numbers in the absence of any further relevant information on such numbers would presumably be the subject of such a merely potential justificatory regress.

(v) *Objection:* “But defenders of infinite justificatory regresses cannot distinguish thus between actual regresses and those that are merely potential. There is no real distinction to be drawn between the two. For if any regress ever justifies the belief at its head, then every regress must always do so. But obviously not every regress does so (as we have seen by examples), and hence no regress can do so.”⁴

Reply: One can in fact distinguish between actual justificatory regresses and merely potential ones, and one can do so both abstractly and by examples.

What an actual regress has that a merely potential regress lacks is the property of containing only justified beliefs as members. What they both share is the property of containing no member without successors that would jointly justify it.

Recall our regress about perfect numbers greater than 100; i.e., there is at least one; there are at least two; there are at least three; and so on. Each member has a successor that would justify it, but no member is justified (in the absence of further information external to the regress). That is therefore a merely potential infinite regress. As for an actual regress, I see no compelling reason why someone (if not a human, then some more powerful mind) could not hold an infinite series of actually justified beliefs as follows:

- (σ2) There is at least one even number
 - There are at least two even numbers
 - There are at least three even numbers

It may be that no one could be the subject of such a series of justified beliefs unless he had a proof that there is a denumerable infinity of even numbers. But even if that should be so, it would not take away the fact of the infinite regress of potential justifiers, each of which is actually justified, and hence it would not take away the fact of the actual endless regress of justification.

The objection under discussion is confused, moreover, on the nature of the issue before us. Our question is *not* whether there can be an infinite potential regress, each member of which would be justified by its successors, such that the belief at its head is justified in virtue of its position there, at the head of such a regress. The existence and even the possibility of a single such regress with a belief at its head that was *not* justified in virtue of its position there would of course settle that question in the negative. Our question is, rather, whether there can be an actual infinite regress of justification, and the fact that a belief at the head of a potential regress might still fail to be justified despite its position does *not* settle this question. For even if there can be a merely potential regress with an unjustified belief at its head, that leaves open the possibility of an infinite regress, each member of which is justified by its immediate successors working jointly, where every member of the regress is in addition actually justified.

6 The Relation of Justification and Foundationalist Strategy

The foregoing discussion is predicated on a simple conception of justification such that a set of beliefs β conditionally justifies (*would* justify) a belief X iff, necessarily, if all members of β are justified then X is also justified (if it exists). The fact that on such a conception of justification actual endless regresses – such as ($\sigma 2$) – seem quite possible blocks a straightforward regress argument in favor of foundations. For it shows that an actual infinite regress cannot be dismissed out of hand.

Perhaps the foundationalist could introduce some relation of justification – presumably more complex and yet to be explicated – with respect to which it could be argued more plausibly that an actual endless regress is out of the question.

There is, however, a more straightforward strategy open to the foundationalist. For he *need not* object to the possibility of an endless regress of justification. His essential creed is the more positive belief that every justified belief must be at the head of a terminating regress. Fortunately, to affirm the universal necessity of a terminating regress is *not* to deny the bare possibility of a nonterminating regress. For a single belief can trail at once regresses of both sorts: one terminating and one not. Thus the proof of the denumerably infinite cardinality of the set of evens may provide for a powerful enough intellect a *terminating* regress for each member of the *endless* series of justified beliefs:

- ($\sigma 2$) There is at least one even number
 There are at least two even numbers
 There are at least three even numbers

At the same time, it is obvious that each member of ($\sigma 2$) lies at the head of an actual endless regress of justification, on the assumption that each member is conditionally justified by its successor, which is in turn actually justified.

“Thank you so much,” the foundationalist may sneer, “but I really do not need that kind of help. Nor do I need to be reminded of my essential creed, which I know as well as anyone. Indeed my rejection of endless regresses of justification is only a means of supporting my view that every justified belief must rest ultimately on foundations, on a terminating regress. You reject that strategy much too casually, in my view, but I will not object here. So we put that strategy aside. And now, my helpful friend, just what do we put in its place?”

Fair enough. How then could one show the need for foundations if an endless regress is not ruled out?

7 Two Levels of Foundationalism

- a. We need to distinguish, first, between two forms of foundationalism: one *formal*, the other *substantive*. A type of *formal foundation-*

alism with respect to a normative or evaluative property ϕ is the view that the conditions (actual and possible) within which ϕ would apply can be specified in general, perhaps recursively. *Substantive foundationalism* is only a particular way of doing so, and coherence is another.

Simpleminded hedonism is the view that:

- (i) every instance of pleasure is good,
- (ii) everything that causes something good is itself good, and
- (iii) everything that is good is so in virtue of (i) or (ii) above.

Simpleminded hedonism is a type of formal foundationalism with respect to the good.

Classical foundationalism in epistemology is the view that:

- (i) every infallible, indubitable belief is justified,
- (ii) every belief deductively inferred from justified beliefs is itself justified, and
- (iii) every belief that is justified is so in virtue of (i) or (ii) above.

Classical foundationalism is a type of formal foundationalism with respect to epistemic justification.

Both of the foregoing theories – simpleminded hedonism in ethics, and classical foundationalism in epistemology – are of course flawed. But they both remain examples of formal foundationalist theories.

- b. One way of arguing in favor of formal foundationalism in epistemology is to formulate a convincing formal foundationalist theory of justification. But classical foundationalism in epistemology no longer has for many the attraction that it had for Descartes, nor has any other form of epistemic foundationalism won general acceptance. Indeed epistemic foundationalism has been generally abandoned, and its advocates have been put on the defensive by the writings of Wittgenstein, Quine, Sellars, Rescher, Aune, Harman, Lehrer, and others. It is lamentable that in our headlong rush away from foundationalism we have lost sight of the different types of foundationalism (formal vs. substantive) and of the different grades of each type. Too many of us now see it as a blur to be decried and avoided. Thus our present attempt to bring it all into better focus.

- c. If we cannot argue from a generally accepted foundationalist theory, what reason is there to accept formal foundationalism? There is no reason to think that the conditions (actual and possible) within which an object is spherical are generally specifiable in nongeometric terms. Why should we think that the conditions (actual and possible) within which a belief is epistemically justified are generally specifiable in nonepistemic terms?

So far as I can see, the main reason for accepting formal foundationalism in the absence of an actual, convincing formal foundationalist theory is the very plausible idea that epistemic justification is subject to the supervenience that characterizes normative and evaluative properties generally. Thus, if a car is a good car, then any physical replica of that car must be just as good. If it is a good car in virtue of such properties as being economical, little prone to break down, etc., then surely any exact replica would share all such properties and would thus be equally good. Similarly, if a belief is epistemically justified, it is presumably so in virtue of its character and its basis in perception, memory, or inference (if any). Thus any belief exactly like it in its character and its basis must be equally well justified. Epistemic justification is supervenient. The justification of a belief supervenes on such properties of it as its content and its basis (if any) in perception, memory, or inference. Such a doctrine of supervenience may itself be considered, with considerable justice, a grade of foundationalism. For it entails that every instance of justified belief is founded on a number of its nonepistemic properties, such as its having a certain basis in perception, memory, and inference, or the like.

But there are higher grades of foundationalism as well. There is, for instance, the doctrine that the conditions (actual and possible) within which a belief would be epistemically justified *can be specified* in general, perhaps recursively (and by reference to such notions as perception, memory, and inference).

A higher grade yet of formal foundationalism requires not only that the conditions for justified belief be specifiable, in general, but that they be specifiable by a simple, comprehensive theory.

- d. Simpleminded hedonism is a formal foundationalist theory of the highest grade. If it is true, then in every possible world goodness supervenes on pleasure and causation in a way that is recursively specifiable by means of a very simple theory.

Classical foundationalism in epistemology is also a formal foundationalist theory of the highest grade. If it is true, then in every possible world epistemic justification supervenes on infallibility cum indubitability and deductive inference in a way that is recursively specifiable by means of a very simple theory.

Surprisingly enough, coherentism may also turn out to be formal foundationalism of the highest grade, provided only that the concept of coherence is itself both simple enough and free of any normative or evaluative admixture. Given these provisos, coherentism explains how epistemic justification supervenes on the nonepistemic in a theory of remarkable simplicity: a belief is justified if it has a place within a system of beliefs that is coherent and comprehensive.

It is a goal of ethics to explain how the ethical rightness of an action supervenes on what is not ethically evaluative or normative. Similarly, it is a goal of epistemology to explain how the epistemic justification of a belief supervenes on what is not epistemically evaluative or normative. If coherentism aims at this goal, that imposes restrictions on the notion of coherence, which must now be conceived innocent of epistemically evaluative or normative admixture. Its substance must therefore consist of such concepts as explanation, probability, and logical implication – with these conceived, in turn, innocent of normative or evaluative content.

- e. We have found a surprising kinship between coherentism and substantive foundationalism, both of which turn out to be varieties of a deeper foundationalism. This deeper foundationalism is applicable to any normative or evaluative property ϕ , and it comes in three grades. The *first* or lowest is simply the supervenience of ϕ : the idea that whenever something has ϕ its having it is founded on certain others of its properties which fall into certain restricted sorts. The *second* is the explicable supervenience of ϕ : the idea that there are formulable principles that explain in quite

general terms the conditions (actual and possible) within which ϕ applies. The *third* and highest is the easily explicable supervenience of ϕ : the idea that there is a *simple* theory that explains the conditions within which ϕ applies. We have found the coherentist and the substantive foundationalist sharing a primary goal: the development of a formal foundationalist theory of the highest grade. For they both want a simple theory that explains precisely how epistemic justification supervenes, in general, on the nonepistemic. This insight gives us an unusual viewpoint on some recent attacks against foundationalism. Let us now consider as an example a certain simple form of argument distilled from the recent antifoundationalist literature.⁵

8 Doxastic Ascent Arguments

Several attacks on foundationalism turn on a sort of “doxastic ascent” argument that calls for closer scrutiny.⁶ Here are two examples:

- A. A belief B is foundationally justified for S in virtue of having property F only if S is justified in believing (1) that most at least of his beliefs with property F are true, and (2) that B has property F. But this means that belief B is not foundational after all, and indeed that the very notion of (empirical) foundational belief is incoherent.

It is sometimes held, for example, that perceptual or observational beliefs are often justified through their origin in the exercise of one or more of our five senses in standard conditions of perception. The advocate of doxastic ascent would raise a vigorous protest, however, for in his view the mere fact of such sensory prompting is impotent to justify the belief prompted. Such prompting must be coupled with the further belief that one’s senses work well in the circumstances, or the like. For we are dealing here with *knowledge*, which requires not blind faith but *reasoned* trust. But now surely the further belief about the reliability of one’s senses itself cannot rest on blind faith but requires its own backing of reasons, and we are off on the regress.

- B. A belief B of proposition P is foundationally justified for S only if S is justified in believ-

ing that there are no factors present that would cause him to make mistakes on the matter of the proposition P. But, again, this means that belief B is not foundational after all and indeed that the notion of (empirical) foundational belief is incoherent.

From the vantage point of formal foundationalism, neither of these arguments seems persuasive. In the first place, as we have seen, what makes a belief foundational (formally) is its having a property that is nonepistemic (not evaluative in the epistemic or cognitive mode), and does not involve inference from other beliefs, but guarantees, via a necessary principle, that the belief in question is justified. A belief B is made foundational by having some such nonepistemic property that yields its justification. Take my belief that I am in pain in a context where it is caused by my being in pain. The property that my belief then has, of being a self-attribution of pain caused by one's own pain is, let us suppose, a nonepistemic property that yields the justification of any belief that has it. So my belief that I am in pain is in that context foundationally justified. Along with my belief that I am in pain, however, there come other beliefs that are equally well justified, such as my belief that someone is in pain. Thus I am foundationally justified in believing that I am in pain only if I am justified in believing that someone is in pain. Those who object to foundationalism as in A or B above are hence mistaken in thinking that their premises would refute foundationalism. The fact is that they would not touch it. For a belief is no less foundationally justified for having its justification yoked to that of another closely related belief.

The advocate of arguments like A and B must apparently strengthen his premises. He must apparently claim that the beliefs whose justification is entailed by the foundationally justified status of belief B must in some sense function as a *necessary source* of the justification of B. And this would of course preclude giving B foundationally justified status. For if the *being justified* of those beliefs is an *essential* part of the source of the justification of B, then it is ruled out that there be a wholly *non-epistemic* source of B's justification.

That brings us to a second point about A and B, for it should now be clear that these cannot be selectively aimed at foundationalism. In particular, they seem neither more nor less valid objections to coherentism than to foundationalism, or so I will now argue about each of them in turn.

A. A belief X is justified for S in virtue of membership in a coherent set only if S is justified in believing (1) that most at least of his beliefs with the property of thus cohering are true, and (2) that X has that property.

Any coherentist who accepts A seems bound to accept A'. For what could he possibly appeal to as a relevant difference? But A' is a quicksand of endless depth. (How is he justified in believing A' (1)? Partly through justified belief that *it* coheres? And what would justify *this*? And so on...)

B'. A belief X is justified for S only if S is justified in believing that there are no factors present that would cause him to make mistakes on the subject matter of that belief.

Again, any coherentist who accepts B seems bound to accept B'. But this is just another road to the quicksand. (For S is justified in believing that there are no such factors only if... and so on.)

Why are such regresses vicious? The key is again, to my mind, the doctrine of supervenience. Such regresses are vicious because they would be logically incompatible with the supervenience of epistemic justification on such nonepistemic facts as the totality of a subject's beliefs, his cognitive and experiential history, and as many other non-epistemic facts as may seem at all relevant. The idea is that there is a set of such nonepistemic facts surrounding a justified belief such that no belief could possibly have been surrounded by those very facts without being justified. Advocates of A or B run afoul of such supervenience, since they are surely committed to the more general views derivable from either A or B by deleting "foundationally" from its first sentence. In each case the more general view would then preclude the possibility of supervenience, since it would entail that the source of justification *always* includes an *epistemic* component.

9 Coherentism and Substantive Foundationalism

a. The notions of coherentism and substantive foundationalism remain unexplicated. We have relied so far on our intuitive grasp of

them. In this section we shall consider reasons for the view that substantive foundationalism is superior to coherentism. To assess these reasons, we need some more explicit account of the difference between the two.

By coherentism we shall mean any view according to which the ultimate sources of justification for any belief lie in relations among that belief and other beliefs of the subject: explanatory relations, perhaps, or relations of probability or logic.

According to substantive foundationalism, as it is to be understood here, there are ultimate sources of justification other than relations among beliefs. Traditionally these additional sources have pertained to the special content of the belief or its special relations to the subjective experience of the believer.

- b. The view that justification is a matter of relations among beliefs is open to an objection from alternative coherent systems or detachment from reality, depending on one's perspective. From the latter perspective the body of beliefs is held constant and the surrounding world is allowed to vary, whereas from the former perspective it is the surrounding world that is held constant while the body of beliefs is allowed to vary. In either case, according to the coherentist, there could be no effect on the justification for any belief.

Let us sharpen the question before us as follows. Is there reason to think that there is at least one system B' , alternative to our actual system of beliefs B , such that B' contains a belief X with the following properties:

- (i) in our present nonbelief circumstances we would not be justified in having belief X even if we accepted along with that belief (as our total system of beliefs) the entire belief system B' in which it is embedded (no matter how acceptance of B' were brought about); and
- (ii) that is so despite the fact that belief X coheres within B' at least as fully as does some actual justified belief of ours within our actual belief system B (where the justification of that actual justified belief is alleged by the coherentist to derive solely from its coherence within our actual body of beliefs B).

The coherentist is vulnerable to counterexamples of this sort right at the surface of his body of beliefs, where we find beliefs with minimal coherence, whose detachment and replacement with contrary beliefs would have little effect on the coherence of the body. Thus take my belief that I have a headache when I do have a splitting headache, and let us suppose that this *does* cohere within my present body of beliefs. (Thus I have no reason to doubt my present introspective beliefs, and so on. And if my belief does *not* cohere, so much the worse for coherentism, since my belief is surely justified.) Here then we have a perfectly justified or warranted belief. And yet such a belief may well have relevant relations of explanation, logic, or probability with at most a small set of other beliefs of mine at the time: say, that I am not free of headache, that I am in pain, that someone is in pain, and the like. If so, then an equally coherent alternative is not far to seek. Let everything remain constant, *including* the splitting headache, except for the following: replace the belief that I have a headache with the belief that I do *not* have a headache, the belief that I am in pain with the belief that I am *not* in pain, the belief that someone is in pain with the belief that someone is *not* in pain, and so on. I contend that my resulting hypothetical system of beliefs would cohere as fully as does my actual system of beliefs, and yet my hypothetical belief that I do *not* have a headache would not therefore be justified. What makes this difference concerning justification between my actual belief that I have a headache and the hypothetical belief that I am free of headache, each as coherent as the other within its own system, if not the actual splitting headache? But the headache is *not* itself a belief nor a relation among beliefs and is thus in no way constitutive of the internal coherence of my body of beliefs.

Some might be tempted to respond by alleging that one's belief about whether or not one has a headache is always *infallible*. But since we could devise similar examples for the various sensory modalities and propositional attitudes, the response given for the case of headache would have to be generalized. In effect, it would have to cover "peripheral" beliefs generally – beliefs at the periphery of one's body of beliefs, minimally coherent with the rest. These peripheral beliefs would all be said to be infallible. That is, again, a possible response, but it leads to a capitulation by the coherentist to the radical foundationalist on a crucial issue that has traditionally divided them: the infallibility of beliefs about one's own subjective states.

What is more, not all peripheral beliefs are about one's own subjective states. The direct realist is probably right that some beliefs about our surroundings are uninferred and yet justified. Consider my present belief that the table before me is oblong. This presumably coheres with such other beliefs of mine as that the table has the same shape as the piece of paper before me, which is oblong, and a different shape than the window frame here, which is square, and so on. So far as I can see, however, there is no insurmountable obstacle to replacing that whole set of coherent beliefs with an equally coherent set as follows: that the table before me is square, that the table has the same shape as the square window frame, and a different shape than the piece of paper, which is oblong, and so on. The important points are (a) that this replacement may be made without changing the rest of one's body of beliefs or any aspect of the world beyond, including one's present visual experience of something oblong, not square, as one looks at the table before one; and (b) that it is so, in part, because of the fact (c) that the subject need not have any beliefs about his present sensory experience.

Some might be tempted to respond by alleging that one's present experience is *self-intimating*, i.e., always necessarily taken note of and reflected in one's beliefs. Thus if anyone has visual experience of something oblong, then he believes that he has such experience. But this would involve a further important concession by the coherentist to the radical foundationalist, who would have been granted two of his most cherished doctrines: the infallibility of introspective belief and the self-intimation of experience.

10 The Foundationalist's Dilemma

The antifoundationalist zeal of recent years has left several forms of foundationalism standing. These all share the conviction that a belief can be justified not only by its coherence within a comprehensive system but also by an appropriate combination of observational content and origin in the use of the senses in standard conditions. What follows presents a dilemma for any foundationalism based on any such idea.

- a. We may surely suppose that beings with observational mechanisms radically unlike ours might also have knowledge of their

environment. (That seems possible even if the radical difference in observational mechanisms precludes overlap in substantive concepts and beliefs.)

- b. Let us suppose that there is such a being, for whom experience of type ϕ (of which we have no notion) has a role with respect to his beliefs of type ϕ analogous to the role that our visual experience has with respect to our visual beliefs. Thus we might have a schema such as that in Table 14.1.

Table 14.1

Human	Extraterrestrial being
Visual experience	ϕ experience
Experience of something red	Experience of something F
Belief that there is something red before one	Belief that there is something F before one

- c. It is often recognized that our visual experience intervenes in two ways with respect to our visual beliefs: as cause and as justification. But these are not wholly independent. Presumably, the justification of the belief that something here is red derives at least in part from the fact that it originates in a visual experience of something red that takes place in normal circumstances.
- d. Analogously, the extraterrestrial belief that something here has the property of being F might be justified partly by the fact that it originates in a ϕ experience of something F that takes place in normal circumstances.
- e. A simple question presents the foundationalist's dilemma: regarding the epistemic principle that underlies our justification for believing that something here is red on the basis of our visual experience of something red, is it proposed as a fundamental principle or as a derived generalization? Let us compare the famous Principle of Utility of value theory, according to which it is best for that to happen which, of all the possible alternatives in the circumstances, would bring with it into the world the greatest balance of pleasure over pain, joy over sorrow, happiness over unhappiness, content

over discontent, or the like. Upon this fundamental principle one may then base various generalizations, rules of thumb, and maxims of public health, nutrition, legislation, etiquette, hygiene, and so on. But these are all then derived generalizations which rest for their validity on the fundamental principle. Similarly, one may also ask, with respect to the generalizations advanced by our foundationalist, whether these are proposed as fundamental principles or as derived maxims or the like. This sets him face to face with a dilemma, each of whose alternatives is problematic. If his proposals are meant to have the status of secondary or derived maxims, for instance, then it would be quite unphilosophical to stop there. Let us turn, therefore, to the other alternative.

- f. On reflection it seems rather unlikely that epistemic principles for the justification of observational beliefs by their origin in sensory experience could have a status more fundamental than that of derived generalizations. For by granting such principles fundamental status we would open the door to a multitude of equally basic principles with no unifying factor. There would be some for vision, some for hearing, etc., without even mentioning the corresponding extraterrestrial principles.
- g. It may appear that there is after all an idea, however, that unifies our multitude of principles. For they all involve sensory experience and sensible characteristics. But what is a sensible characteristic? Aristotle's answer appeals to examples: colors, shapes, sounds, and so on. Such a notion might enable us to unify perceptual epistemic principles under some more fundamental principle such as the following.

If σ is a sensible characteristic, then the belief that there is something with σ before one is (prima facie) justified if it is based on a visual experience of something with σ in conditions that are normal with respect to σ .

- h. There are at least two difficulties with such a suggestion, however, and neither one can be brushed aside easily. First, it is not clear that we can have a viable notion of sensible

characteristics on the basis of examples so diverse as colors, shapes, tones, odors, and so on. Second, the authority of such a principle apparently derives from contingent circumstances concerning the reliability of beliefs prompted by sensory experiences of certain sorts. According to the foundationalist, our visual beliefs are justified by their origin in our visual experience or the like. Would such beliefs be equally well justified in a world where beliefs with such an origin were nearly always false?

- i. In addition, finally, even if we had a viable notion of such characteristics, it is not obvious that fundamental knowledge of reality would have to derive causally or otherwise from sensory experience of such characteristics. How could one impose reasonable limits on extraterrestrial mechanisms for noninferential acquisition of beliefs? Is it not possible that such mechanisms need not always function through sensory experience of any sort? Would such beings necessarily be denied any knowledge of the surroundings and indeed of any contingent spatio-temporal fact? Let us suppose them to possess a complex system of true beliefs concerning their surroundings, the structures below the surface of things, exact details of history and geography, all constituted by concepts none of which corresponds to any of our sensible characteristics. What then? Is it not possible that their basic beliefs should all concern fields of force, waves, mathematical structures, and numerical assignments to variables in several dimensions? This is no doubt an exotic notion, but even so it still seems conceivable. And if it is in fact possible, what then shall we say of the noninferential beliefs of such beings? Would we have to concede the existence of special epistemic principles that can validate their noninferential beliefs? Would it not be preferable to formulate more abstract principles that can cover both human and extraterrestrial foundations? If such more abstract principles are in fact accessible, then the less general principles that define the human foundations and those that define the extraterrestrial foundations are both derived principles whose validity depends on that of the more abstract principles. In this the human and extraterrestrial epistemic principles would

resemble rules of good nutrition for an infant and an adult. The infant's rules would of course be quite unlike those valid for the adult. But both would still be based on a more fundamental principle that postulates the ends of well-being and good health. What more fundamental principles might support both human and extraterrestrial knowledge in the way that those concerning good health and well-being support rules of nutrition for both the infant and adult?

11 Reliabilism: An Ethics of Moral Virtues and an Epistemology of Intellectual Virtues

In what sense is the doctor attending Frau Hitler justified in performing an action that brings with it far less value than one of its accessible alternatives? According to one promising idea, the key is to be found in the rules that he embodies through stable dispositions. His action is the result of certain stable virtues, and there are no equally virtuous alternative *dispositions* that, given his cognitive limitations, he might have embodied with equal or better total consequences, and that would have led him to infanticide in the circumstances. The important move for our purpose is the stratification of justification. Primary justification attaches to virtues and other dispositions, to stable dispositions to act, through their greater contribution of value when compared with alternatives. Secondary justification attaches to particular acts in virtue of their source in virtues or other such justified dispositions.

The same strategy may also prove fruitful in epistemology. Here primary justification would apply to *intellectual* virtues, to stable dispositions for belief acquisition, through their greater contribution toward getting us to the truth. Secondary justification would then attach to particular beliefs in virtue of their source in intellectual virtues or other such justified dispositions.⁷

That raises parallel questions for ethics and epistemology. We need to consider more carefully the concept of a virtue and the distinction between moral and intellectual virtues. In epistemology, there is reason to think that the most useful and illuminating notion of intellectual virtue will prove broader than our tradition would suggest and must

give due weight not only to the subject and his intrinsic nature but also to his environment and to his epistemic community. This is a large topic, however, to which I hope some of us will turn with more space, and insight, than I can now command.

Summary

1. *Two assumptions*: (A1) that for a belief to constitute knowledge it must be (a) true and (b) justified; and (A2) that the justification relevant to whether or not one knows is a sort of epistemic or theoretical justification to be distinguished from its practical counterpart.
2. *Knowledge and criteria*. Particularism is distinguished from methodism: the first gives priority to particular examples of knowledge over general methods or criteria, whereas the second reverses that order. The methodism of Descartes leads him to an elaborate dogmatism whereas that of Hume leads him to a very simple skepticism. The particularist is, of course, antiskeptical on principle.
3. *Two metaphors: the raft and the pyramid*. For the foundationalist every piece of knowledge stands at the apex of a pyramid that rests on stable and secure foundations whose stability and security do not derive from the upper stories or sections. For the coherentist a body of knowledge is a free-floating raft every plank of which helps directly or indirectly to keep all the others in place, and no plank of which would retain its status with no help from the others.
4. *A coherentist critique of foundationalism*. No mental state can provide a foundation for empirical knowledge. For if such a state is propositional, then it is fallible and hence no secure foundation. But if it is *not* propositional, then how can it possibly serve as a foundation for belief? How can one infer or justify anything on the basis of a state that, having no propositional content, must be logically dumb? An analogy with ethics suggests a reason to reject this dilemma. Other reasons are also advanced and discussed.
5. *The regress argument*. In defending his position, the foundationalist often attempts to rule out the very possibility of an infinite regress of justification (which leads him to

- the necessity for a foundation). Some of his arguments to that end are examined.
6. *The relation of justification and foundationalist strategy.* An alternative foundationalist strategy is exposed, one that does not require ruling out the possibility of an infinite regress of justification.
 7. *Two levels of foundationalism.* Substantive foundationalism is distinguished from formal foundationalism, three grades of which are exposed: first, the supervenience of epistemic justification; second, its explicable supervenience; and, third, its supervenience explicable by means of a simple theory. There turns out to be a surprising kinship between coherentism and substantive foundationalism, both of which aim at a formal foundationalism of the highest grade, at a theory of the greatest simplicity that explains how epistemic justification supervenes on nonepistemic factors.
 8. *Doxastic ascent arguments.* The distinction between formal and substantive foundationalism provides an unusual viewpoint on some recent attacks against foundationalism. We consider doxastic ascent arguments as an example.
 9. *Coherentism and substantive foundationalism.* It is argued that substantive foundationalism is superior, since coherentism is unable to account adequately for the epistemic status of beliefs at the "periphery" of a body of beliefs.
 10. *The foundationalist's dilemma.* All foundationalism based on sense experience is subject to a fatal dilemma.
 11. *Reliabilism.* An alternative to foundationalism of sense experience is sketched.

Appendix⁸

What one is rationally justified in believing obviously depends on the data in one's possession. But what data one has can depend on how much and how well one investigates. Consider, therefore, the following possibility. What if A is rationally justified in believing x given his body of data D_1 whereas B is not rationally justified in believing x given his body of data D_2 , where D_2 includes D_1 but is much more extensive as a result of A's irresponsible negligence and B's commendable thoroughness? The present account might unfortunately grant A knowledge while denying it to B,

for A's neglect so far has no bearing on any epistemic pyramid.

We have considered a situation where someone lacks knowledge owing to his misuse of his cognitive equipment, either by letting it idle when it should be functioning or by busily employing it dysfunctionally. Another situation where someone lacks knowledge despite having rationally justified correct belief might be called the Magoo situation – where S lacks adequate equipment to begin with (relative to the question in hand: whether p).⁹ It is because of this type of lack that despite his extensive experience with cable cars, Mr Magoo does not know that his cable car will arrive safely when, unknown to him, bombs are raining all around it. Of course, even if you have less than 20–20 vision you can still know that there is an elephant in front of you when you see one there. So not just any defect will make your equipment inadequate for a judgment on the question whether p . I would venture that it must be a defect that prevents you from acquiring information that (i) a normal inquirer in the epistemic community would acquire in that situation *and* (ii) makes a difference to what you can reasonably conclude on the question whether p (or at least to how reasonably you can draw the conclusion).

The possibility of inadequate cognitive equipment requires a further and more striking departure from the traditional conception of knowledge. Despite having warranted correct belief, someone may lack knowledge owing to his neglectful data-collection. There lack of knowledge could be traced back to epistemic irresponsibility, to substandard performance blamed on the investigator. In the present example, blame is out of place. By hypothesis, Magoo conducts impeccable "inquiry" both in arriving at his data and on the basis of his data. But he still falls short of knowledge, despite his warranted, correct belief. His shortcoming is substandard equipment, for which we may suppose him to be blameless. Hence something other than epistemic justification or correct belief can help determine what one knows or does not know. Even if one correctly believes that p with full rational justification and free of irrational or neglectful unbelief, one may still be in no position to know, because of faulty cognitive equipment.

In all of the foregoing cases, someone misses or is liable to miss available information which may be highly relevant and important and may make a difference to what he can conclude on the question

in hand. In each case, moreover, he seems culpable or discredited in some sense: he would seem less reliable than otherwise for his role in any such case. But there appear to be situations where again someone misses available information with no culpability *or* discredit. Harman gives an example where S reads in a newspaper that some famous person has been assassinated, but does not read the next edition, where all reports of the assassination are denied by highly authoritative and trustworthy people. If practically the whole country reads the next edition and people don't know what to believe, does S alone know of the assassination, provided the next edition is in fact a pack of lies?¹⁰ I suppose we would be inclined to say that he does not know (especially if had he read the next edition, *he* would not have known what to believe). But what if only two or three people get a chance to read the next edition before it is recalled by the newspaper? Should we now say that out of millions who read the first story and mourn the loved leader not one knows of his death? I suppose we would be inclined to say that the fake edition and the few deceived by it make no difference concerning what everybody else knows. It seems plausible to conclude that knowledge has a further "social aspect," that it cannot depend on one's missing or blinking what is generally known.

Our departures from the traditional conception of knowledge put in relief the relativity of knowledge to an epistemic community. This is brought out most prominently by the requirement that inquirers have at least *normal* cognitive equipment (e.g., normal perceptual apparatus, where that is relevant). But our new requirement – that inquirers not lack or blink generally known relevant information – also brings out the relativity. A vacationer in the woods may know that *p* well enough for an average vacationer, but he won't have the kind of knowledge his guide has. A guide would scornfully deny that the tenderfoot really knows that *p*. Relative to the epistemic community of guides (for that area) the tenderfoot lacks relevant generally known information, and misses relevant data that the average guide would grasp in the circumstances.

These departures from the traditional account may make better sense if we reflect that the honorific term "knowledgeable" is to be applied only to those who are reliable sources of information, surely an important category for a language-using, social species.

We have now taken note of two types of situation where correct, fully warranted belief falls short of knowledge owing to no neglect or faulty reasoning or false belief. Despite commendable thoroughness and impeccable reasoning unspoiled by falsehood, one may still fail to be "in a position to know," owing either to faulty cognitive equipment or to missed generally known information. I am not suggesting that these are the only ways to be out of position to know. I have no complete list of epistemic principles describing ways of arriving at a position to know or of being blocked from such a position. My suggestion is only that there are such principles, and that in any case we must go beyond the traditional emphasis by epistemologists on warrant and reasoning as determinants of knowledge. Despite the importance of warranted correct belief in determining what we know, the Gettier examples show that it is not alone enough to guarantee knowledge. What is more, warranted correct belief supported by reasoning *unspoiled by falsehood* seems immune to Gettier examples, but it still falls short of knowledge, as we have seen.

My conclusion is that to understand knowledge we must enrich our traditional repertoire of epistemic concepts with the notion of *being in a position to know* (from the *point of view of a K*, e.g., a human being). Thus a proposition is evident (from the point of view of a *K*) to a subject only if *both* he is rationally justified in believing it *and* he is in a position to know (from the *K* point of view) whether it is true. It may *be* (and not just appear) evident to Magoo from *his* point of view that he will reach the other side safely, but it seems wrong to say of Magoo as he steps into the cable car with bombs raining all around that it *is* quite evident to him that he will arrive safely. It seems wrong *for whom* to say this? For one of us, naturally; that is, for a normal human from *his* point of view. And since a normal human could not help seeing and hearing the bombs, from the human point of view Magoo is not in a position to know that he will arrive safely, inasmuch as he is missing relevant information that a normal human would gather in the circumstances. Hence Magoo does not have *human* knowledge that he will arrive safely, for it is not evident to him from the human point of view that he will so arrive.

Consider this account:

- (A) S knows that *p* iff
 - (a) it is true that *p*;

- (b) S believes that p ; and
- (c) there is a non-defective epistemic pyramid for S and the proposition that p .

Every node of such a pyramid must be true *and* evident. And for every node n that has successors, the successors must serve as grounds that give the subject S rational warrant for believing n . What now seems too narrow about this account emerges with the explanation of what a pyramid of knowledge is, and of what the evident is. For in this explanation what is evident to S is identified with what S is rationally justified in believing. But it now seems plain that for x to be evident to S, *two* conditions must be satisfied: (i) that S be rationally justified in believing x , and (ii) that S be in a position to know whether x is true. And we must also take note of the relativity of knowledge to an epistemic community. Let us therefore replace (A) with the following:

- (B) S knows (from the K point of view) that p iff
 - (a) it is true that p ;
 - (b) S believes that p ; and
 - (c) there is a non-defective epistemic pyramid (from the K point of view) for S and the proposition that p .

Every node of such a pyramid must now be true and evident from the K point of view.

Normally when epistemologists discuss knowledge (of the colors and shapes of surrounding objects, of one's own or one's neighbor's mental states, and so on), they plainly do so from the *human* point of view. But other points of view are possible even in ordinary conversation. The expert/layman distinction is replicable in many different contexts, and with each replication we have a new epistemically relevant distinction in points of view, with expert knowledge on one side and layman knowledge on the other.

Neither Magoo nor the newspaper reader who alone has not seen the new edition is in a position to know (from the human point of view) about the relevant subject matter. Thus we can understand their ignorance and, by parity of reasoning, the ignorance of all those who are out of position to know that p because they lack either adequate cognitive equipment or relevant information that is generally known to those who have taken an epistemic stand on the question whether p (where to suspend judgment *is* to take an epistemic stand, whereas to be totally oblivious to the matter is not).

What it is for S's belief that p to be fully grounded has been explained by means of our epistemic pyramids. That answer points in the right direction, but it can be made more precise: e.g., by clarifying the grounding relation. Moreover, we have found that a fully grounded correct belief is not necessarily knowledge, and this for at least two reasons: (i) it may rest directly or indirectly on some false ground, and (ii) the believer may not be in a position to know.

We have tried to allow for these possibilities by broadening epistemic pyramids, by making room for our new epistemic notion of being-in-a-position-to-know, and by noting that to support knowledge epistemic pyramids must be non-defective, i.e., must contain no false nodes. But pyramids are objectionable for other reasons as well: (i) they may mislead by suggesting that terminal nodes provide a "foundation" in one or another undesirable sense, or by suggesting that terminal nodes must come first in time, so that one may later build on them; (ii) more seriously, there is an unacceptable vagueness in the very idea of such a pyramid, which derives mainly from the vagueness of the "grounding" relation in terms of which pyramids were defined. What follows is an attempt to solve these problems by switching pyramids upside down into trees.

Let us emphasize, however, that this will not commit one to a picture of knowledge according to which there is a bedrock of self-evident propositions. It is perfectly consistent with the present theory that part of what makes *any* proposition evident is its coherence with a network of mutually supporting propositions. Since there is bound to be a multitude of such coherent networks, however, a non-arbitrary narrowing of the field must be supported by something other than coherence.

We turn finally to an account (C) according to which S knows that p provided that both (a) S correctly believes that p ,¹¹ and (b) there is a set of propositions that fully and non-defectively renders it evident to S that p (where a set "non-defectively renders it evident to S that p " if and only if it does so without attributing to S any false belief).¹²

Supposing this account correct, every bit of knowledge has a tree like that shown in Figure 14.3. Note that each node of such a tree is a proposition. Thus the "root" node is the-proposition-that- p_1 , and the first terminal node (from the left) the- proposition-that- p_{111} .¹³

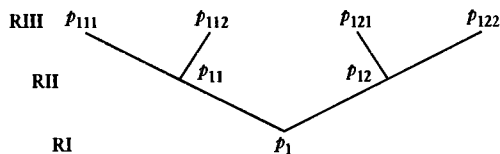


Figure 14.3

There is an important difference between these trees and our earlier pyramids. Except for terminal nodes, every node of a tree is an epistemic proposition, whereas not a single node of a pyramid need be epistemic at all. Pyramids display propositions that are evident to A (not propositions that such and such other propositions are evident to S), and they also show which propositions ground (for S) any proposition for which S has grounds. Trees

display true epistemic propositions concerning S and they also show what “makes these propositions true” *via* epistemic principles. A tree must do this for *every* epistemic proposition that constitutes one of its nodes. That is to say, trees contain no epistemic terminal nodes. It is in this sense that trees provide *complete* epistemic explanations of the truth of their root nodes.

Notes

- 1 But Descartes's methodism was at most partial. James Van Cleve has supplied the materials for a convincing argument that the way out of the Cartesian circle is through a particularism of basic knowledge. See James Van Cleve, “Foundationalism, Epistemic Principles, and the Cartesian Circle,” this vol., ch. 20. But this is, of course, compatible with methodism on inferred knowledge. Whether Descartes subscribed to such methodism is hard (perhaps impossible) to determine, since in the end he makes room for all the kinds of knowledge required by particularism. But his language when he introduces the method of hyperbolic doubt, and the order in which he proceeds, suggest that he did subscribe to such methodism.
- 2 Cf. Laurence Bonjour, “The Coherence Theory of Empirical Knowledge,” *Philosophical Studies* 30, pp. 281–312, and, especially, Michael Williams, *Groundless Belief* (Oxford: Blackwell, 1977); and Bonjour, “Can Empirical Knowledge Have a Foundation?,” this vol., ch. 21.
- 3 Cf. Richard Foley, “Inferential Justification and the Infinite Regress,” *American Philosophical Quarterly* 15 (1978), pp. 311–16.
- 4 Cf. John Post, “Infinite Regresses of Justification and of Explanation,” *Philosophical Studies* 34 (1980).
- 5 The argument of this whole section is developed in greater detail in my paper “The Foundations of Foundationalism,” *Nous* 14, pp. 547–65.
- 6 For some examples of the influence of doxastic ascent arguments, see Wilfrid Sellars's writing in epistemology: “Empiricism and the Philosophy of Mind,” in *Science, Perception and Reality* (London: Routledge and Kegan Paul, 1963), esp. section VIII, and particularly p. 168. Also I. T. Oakley, “An Argument for Skepticism Concerning Justified Belief,” *American*

Philosophical Quarterly 13 (1976), pp. 221–8; and Bonjour, “Can Empirical Knowledge Have a Foundation?”

- 7 This puts in a more traditional perspective the contemporary effort to develop a “causal theory of knowing.” From our viewpoint, this effort is better understood not as an attempt to *define* propositional knowledge, but as an attempt to formulate fundamental principles of justification.

Cf. the work of D. Armstrong, *Belief, Truth and Knowledge* (Cambridge: Cambridge University Press, 1973), and that of F. Dretske, A. Goldman, and M. Swain, whose relevant already published work is included in G. Pappas and M. Swain (eds), *Essays on Knowledge and Justification* (Ithaca and London, 1978). But the theory is still under development by Goldman and by Swain, who have reached general conclusions about it similar to those suggested here, though not necessarily – so far as I know – for the same reasons or in the same overall context.

- 8 From “How Do You Know?” *American Philosophical Quarterly* 11, 2 (1974), pp. 113–22.
- 9 The Magoo situation is the situation of that unfortunate nearsighted and hearing-impaired cartoon character who fortunately escapes disaster at every turn.
- 10 Gilbert Harman, “Induction,” in Marshall Swain (ed.), *Induction, Acceptance and Rational Belief* (Dordrecht: Reidel, 1970), esp. Sect. IV, pp. 95–7.
- 11 Whether knowledge entails belief at all is of course a vexed question of long standing, but there is no room for it here. A helpful and interesting discussion is found in Keith Lehrer's “Belief and Knowledge,” *Philosophical Review* 77 (1968), pp. 491–9.

- 12 In what follows, the relativity of knowledge to an epistemic community is left implicit, as it normally is in ordinary thought and speech.
- 13 Strictly speaking, what we have here is obviously a *partial tree schema*. For convenience, however, I

speak of trees even when I mean partial tree schemata. Also, it should not be thought that every tree must have exactly three ranks (RI, RII, and RIII). On the contrary, a tree may have any number of ranks, so long as it has more than one.

A Coherence Theory of Truth and Knowledge

Donald Davidson

In this paper I defend what may as well be called a coherence theory of truth and knowledge. The theory I defend is not in competition with a correspondence theory, but depends for its defense on an argument that purports to show that coherence yields correspondence.

The importance of the theme is obvious. If coherence is a test of truth, there is a direct connection with epistemology, for we have reason to believe many of our beliefs cohere with many others, and in that case we have reason to believe many of our beliefs are true. When the beliefs are true, then the primary conditions for knowledge would seem to be satisfied.

Someone might try to defend a coherence theory of truth without defending a coherence theory of knowledge, perhaps on the ground that the holder of a coherent set of beliefs might lack a reason to believe his beliefs coherent. This is not likely, but it may be that someone, though he has true beliefs, and good reasons for holding them, does not appreciate the relevance of reason to belief. Such a one may best be viewed as having knowledge he does not know he has: he thinks he is a skeptic. In a word, he is a philosopher.

Setting aside aberrant cases, what brings truth and knowledge together is meaning. If meanings are given by objective truth conditions there is a question how we can know that the conditions are satisfied, for this would appear to require a confrontation between what we believe and reality; and the idea of such a confrontation is absurd.

Originally published in Ernest LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (New York: Blackwell, 1989), pp. 307-19.

But if coherence is a test of truth, then coherence is a test for judging that objective truth conditions are satisfied, and we no longer need to explain meaning on the basis of possible confrontation. My slogan is: correspondence without confrontation. Given a correct epistemology, we can be realists in all departments. We can accept objective truth conditions as the key to meaning, a realist view of truth, and we can insist that knowledge is of an objective world independent of our thought or language.

Since there is not, as far as I know, a theory that deserves to be called 'the' coherence theory, let me characterize the sort of view I want to defend. It is obvious that not every consistent set of interpreted sentences contains only true sentences, since one such set might contain just the consistent sentence *S* and another just the negation of *S*. And adding more sentences, while maintaining consistency, will not help. We can imagine endless stat-descriptions — maximal consistent descriptions which do not describe our world.

My coherence theory concerns beliefs, or sentences held true by someone who understands them. I do not want to say, at this point, that every possible coherent set of beliefs is true (or contains mostly true beliefs). I shy away from this because it is so unclear what is possible. At one extreme, it might be held that the range of possible maximal sets of beliefs is as wide as the range of possible maximal sets of sentences, and then there would be no point to insisting that a defensible coherence theory concerns beliefs and not propositions or sentences. But there are other ways of conceiving what it is possible to believe which would justify saying not only that all actual

coherent belief systems are largely correct but that all possible ones are also. The difference between the two notions of what it is possible to believe depends on what we suppose about the nature of belief, its interpretation, its causes, its holders, and its patterns. Beliefs for me are states of people with intentions, desires, sense organs; they are states that are caused by, and cause, events inside and outside the bodies of their entertainers. But even given all these constraints, there are many things people do believe, and many more that they could. For all such cases, the coherence theory applies.

Of course some beliefs are false. Much of the point of the concept of belief is the potential gap it introduces between what is held to be true and what is true. So mere coherence, no matter how strongly coherence is plausibly defined, can not guarantee that what is believed is so. All that a coherence theory can maintain is that most of the beliefs in a coherent total set of beliefs are true.

This way of stating the position can at best be taken as a hint, since there is probably no useful way to count beliefs, and so no clear meaning to the idea that most of a person's beliefs are true. A somewhat better way to put the point is to say there is a presumption in favor of the truth of a belief that coheres with a significant mass of belief. Every belief in a coherent total set of beliefs is justified in the light of this presumption, much as every intentional action taken by a rational agent (one whose choices, beliefs and desires cohere in the sense of Bayesian decision theory) is justified. So to repeat, if knowledge is justified true belief, then it would seem that all the true beliefs of a consistent believer constitute knowledge. This conclusion, though too vague and hasty to be right, contains an important core of truth, as I shall argue. Meanwhile I merely note the many problems asking for treatment: what exactly does coherence demand? How much of inductive practice should be included, how much of the true theory (if there is one) of evidential support must be in there? Since no person has a completely consistent body of convictions, coherence with *which* beliefs creates a presumption of truth? Some of these problems will be put in better perspective as I go along.

It should be clear that I do not hope to define truth in terms of coherence and belief. Truth is beautifully transparent compared to belief and coherence, and I take it as primitive. Truth, as applied to utterances of sentences, shows the quotational feature enshrined in Tarski's Conven-

tion T, and that is enough to fix its domain of application. Relative to a language or a speaker, of course, so there is more to truth than Convention T; there is whatever carries over from language to language or speaker to speaker. What Convention T, and the trite sentences it declares true, like "Grass is green" spoken by an English speaker, is true if and only if grass is green, reveal is that the truth of an utterance depends on just two things: what the words as spoken mean, and how the world is arranged. There is no further relativism to a conceptual scheme, a way of viewing things, a perspective. Two interpreters, as unlike in culture, language and point of view as you please, can disagree over whether an utterance is true, but only if they differ on how things are in the world they share, or what the utterance means.

I think we can draw two conclusions from these simple reflections. First, truth is correspondence with the way things are. (There is no straightforward and non-misleading way to state this; to get things right, a detour is necessary through the concept of satisfaction in terms of which truth is characterized.¹) So if a coherence theory of truth is acceptable, it must be consistent with a correspondence theory. Second, a theory of knowledge that allows that we can know the truth must be a non-relativized, non-internal form of realism. So if a coherence theory of knowledge is acceptable, it must be consistent with such a form of realism. My form of realism seems to be neither Hilary Putnam's internal realism nor his metaphysical realism.² It is not internal realism because internal realism makes truth relative to a scheme, and this is an idea I do not think is intelligible.³ A major reason, in fact, for accepting a coherence theory is the unintelligibility of the dualism of a conceptual scheme and a 'world' waiting to be coped with. But my realism is certainly not Putnam's metaphysical realism, for it is characterized by being 'radically non-epistemic', which implies that all our best researched and established thoughts and theories may be false. I think the independence of belief and truth requires only that *each* of our beliefs may be false. But of course a coherence theory cannot allow that all of them can be wrong.

But why not? Perhaps it is obvious that the coherence of a belief with a substantial body of belief enhances its chance of being true, provided there is reason to suppose the body of belief is true, or largely so. But how can coherence alone supply grounds for belief? Mayhap the best we can do to

justify one belief is to appeal to other beliefs. But then the outcome would seem to be that we must accept philosophical skepticism, no matter how unshaken in practice our beliefs remain.

This is skepticism in one of its traditional garbs. It asks: Why couldn't all my beliefs hang together and yet be comprehensively false about the actual world? Mere recognition of the fact that it is absurd or worse to try to *confront* our beliefs, one by one, or as a whole, with what they are about does not answer the question nor show the question unintelligible. In short, even a mild coherence theory like mine must provide a skeptic with a reason for supposing coherent beliefs are true. The partisan of a coherence theory can't allow assurance to come from outside the system of belief, while nothing inside can produce support except as it can be shown to rest, finally or at once, on something independently trustworthy.

It is natural to distinguish coherence theories from others by reference to the question whether or not justification can or must come to an end. But this does not define the positions, it merely suggests a form the argument may take. For there are coherence theorists who hold that some beliefs can serve as the basis for the rest, while it would be possible to maintain that coherence is not enough, although giving reasons never comes to an end. What distinguishes a coherence theory is simply the claim that nothing can count as a reason for holding a belief except another belief. Its partisan rejects as unintelligible the request for a ground or source of justification of another ilk. As Rorty⁷ has put it, 'nothing counts as justification unless by reference to what we already accept, and there is no way to get outside our beliefs and our language so as to find some test other than coherence.'⁴ About this I am, as you see, in agreement with Rorty. Where we differ, if we do, is on whether there remains a question how, given that we cannot 'get outside our beliefs and our language so as to find some test other than coherence', we nevertheless can have knowledge of, and talk about, an objective public world which is not of our own making. I think this question does remain, while I suspect that Rorty doesn't think so. If this is his view, then he must think I am making a mistake in trying to answer the question. Nevertheless, here goes.

It will promote matters at this point to review very hastily some of the reasons for abandoning the search for a basis for knowledge outside the scope of our beliefs. By 'basis' here I mean

specifically an epistemological basis, a source of justification.

The attempts worth taking seriously attempt to ground belief in one way or another on the testimony of the senses: sensation, perception, the given, experience, sense data, the passing show. All such theories must explain at least these two things: what, exactly, is the relation between sensation and belief that allows the first to justify the second? and, why should we believe our sensations are reliable, that is, why should we trust our senses?

The simplest idea is to identify certain beliefs with sensations. Thus Hume seems not to have distinguished between perceiving a green spot and perceiving that a spot is green. (An ambiguity in the word 'idea' was a great help here.) Other philosophers noted Hume's confusion, but tried to attain the same results by reducing the gap between perception and judgement to zero by attempting to formulate judgements that do not go beyond stating that the perception or sensation or presentation exists (whatever that may mean). Such theories do not justify beliefs on the basis of sensations, but try to justify certain beliefs by claiming that they have exactly the same epistemic content as a sensation. There are two difficulties with such a view: first, if the basic beliefs do not exceed in content the corresponding sensation they cannot support any inference to an objective world; and second, there are no such beliefs.

A more plausible line is to claim that we cannot be wrong about how things appear to us to be. If we believe we have a sensation, we do; this is held to be an analytic truth, or a fact about how language is used.

It is difficult to explain this supposed connection between sensations and some beliefs in a way that does not invite skepticism about other minds, and in the absence of an adequate explanation, there should be a doubt about the implications of the connection for justification. But in any case, it is unclear how, on this line, sensations justify the belief in those sensations. The point is rather that such beliefs require no justification, for the existence of the belief entails the existence of the sensation, and so the existence of the belief entails its own truth. Unless something further is added, we are back to another form of coherence theory.

Emphasis on sensation or perception in matters epistemological springs from the obvious thought: sensations are what connect the world and our beliefs, and they are candidates for justifiers

because we often are aware of them. The trouble we have been running into is that the justification seems to depend on the awareness, which is just another belief.

Let us try a bolder tack. Suppose we say that sensations themselves, verbalized or not, justify certain beliefs that go beyond what is given in sensation. So, under certain conditions, having the sensation of seeing a green light flashing may justify the belief that a green light is flashing. The problem is to see how the sensation justifies the belief. Of course if someone has the sensation of seeing a green light flashing, it is likely, under certain circumstances, that a green light is flashing. *We* can say this, since we know of his sensation, but *he* can't say it, since we are supposing he is justified without having to depend on believing he has the sensation. Suppose he believed he didn't have the sensation. Would the sensation still justify him in the belief in an objective flashing green light?

The relation between a sensation and a belief cannot be logical, since sensations are not beliefs or other propositional attitudes. What then is the relation? The answer is, I think, obvious: the relation is causal. Sensations cause some beliefs and in *this* sense are the basis or ground of those beliefs. But a causal explanation of a belief does not show how or why the belief is justified.

The difficulty of transmuting a cause into a reason plagues the anticoherentist again if he tries to answer our second question: What justifies the belief that our senses do not systematically deceive us? For even if sensations justify belief in sensation, we do not yet see how they justify belief in external events and objects.

Quine tells us that science tells us that 'our only source of information about the external world is through the impact of light rays and molecules upon our sensory surfaces'.⁵ What worries me is how to read the words 'source' and 'information'. Certainly it is true that events and objects in the external world cause us to believe things about the external world, and much, if not all, of the causality takes a route through the sense organs. The notion of information, however, applies in a non-metaphorical way only to the engendered beliefs. So 'source' has to be read simply as 'cause' and 'information' as 'true belief' or 'knowledge'. Justification of beliefs caused by our senses is not yet in sight.⁶

The approach to the problem of justification we have been tracing must be wrong. We have been

trying to see it this way: a person has all his beliefs about the world – that is, all his beliefs. How can he tell if they are true, or apt to be true? Only, we have been assuming, by connecting his beliefs to the world, confronting certain of his beliefs with the deliverances of the senses one by one, or perhaps confronting the totality of his beliefs with the tribunal of experience. No such confrontation makes sense, for of course we can't get outside our skins to find out what is causing the internal happenings of which we are aware. Introducing intermediate steps or entities into the causal chain, like sensations or observations, serves only to make the epistemological problem more obvious. For if the intermediaries are merely causes, they don't justify the beliefs they cause, while if they deliver information, they may be lying. The moral is obvious. Since we can't swear intermediaries to truthfulness, we should allow no intermediaries between our beliefs and their objects in the world. Of course there are causal intermediaries. What we must guard against are epistemic intermediaries.

There are common views of language that encourage bad epistemology. This is no accident, of course, since theories of meaning are connected with epistemology through attempts to answer the question how one determines that a sentence is true. If knowing the meaning of a sentence (knowing how to give a correct interpretation of it) involves, or is, knowing how it could be recognized to be true, then the theory of meaning raises the same question we have been struggling with, for giving the meaning of a sentence will demand that we specify what would justify asserting it. Here the coherentist will hold that there is no use looking for a source of justification outside of other sentences held true, while the foundationalist will seek to anchor at least some words or sentences to non-verbal rocks. This view is held, I think, both by Quine and by Michael Dummett.

Dummett and Quine differ, to be sure. In particular, they disagree about holism, the claim that the truth of our sentences must be tested together rather than one by one. And they disagree also, and consequently, about whether there is a useful distinction between analytic and synthetic sentences, and about whether a satisfactory theory of meaning can allow the sort of indeterminacy Quine argues for. (On all these points, I am Quine's faithful student.)

But what concerns me here is that Quine and Dummett agree on a basic principle, which is that

whatever there is to meaning must be traced back somehow to experience, the given, or patterns of sensory stimulation, something intermediate between belief and the usual objects our beliefs are about. Once we take this step, we open the door to skepticism, for we must then allow that a very great many – perhaps most – of the sentences we hold to be true may in fact be false. It is ironical. Trying to make meaning accessible has made truth inaccessible. When meaning goes epistemological in this way, truth and meaning are necessarily divorced. One can, of course, arrange a shotgun wedding by redefining truth as what we are justified in asserting. But this does not marry the original mates.

Take Quine's proposal that whatever there is to the meaning (information value) of an observation sentence is determined by the patterns of sensory stimulation that would cause a speaker to assent to or dissent from the sentence. This is a marvelously ingenious way of capturing what is appealing about verificationist theories without having to talk of meanings, sense-data, or sensations; for the first time it made plausible the idea that one could, and should, do what I call the theory of meaning without need of what Quine calls meanings. But Quine's proposal, like other forms of verificationism, makes for skepticism. For clearly a person's sensory stimulations could be just as they are and yet the world outside very different. (Remember the brain in the vat.)

Quine's way of doing without meanings is subtle and complicated. He ties the meanings of some sentences directly to patterns of stimulation (which also constitute the evidence, Quine thinks, for assenting to the sentence), but the meanings of further sentences are determined by how they are conditioned to the original, or observation sentences. The facts of such conditioning do not permit a sharp division between sentences held true by virtue of meaning and sentences held true on the basis of observation. Quine made this point by showing that if one way of interpreting a speaker's utterances was satisfactory, so were many others. This doctrine of the indeterminacy of translation, as Quine called it, should be viewed as neither mysterious nor threatening. It is no more mysterious than the fact that temperature can be measured in Centigrade or Fahrenheit (or any linear transformation of those numbers). And it is not threatening because the very procedure that demonstrates the degree of indeterminacy at the same time

demonstrates that what is determinate is all we need.

In my view, erasing the line between the analytic and synthetic saved philosophy of language as a serious subject by showing how it could be pursued without what there cannot be: determinate meanings. I now suggest also giving up the distinction between observation sentences and the rest. For the distinction between sentences belief in whose truth is justified by sensations and sentences belief in whose truth is justified only by appeal to other sentences held true is as anathema to the coherentist as the distinction between beliefs justified by sensations and beliefs justified only by appeal to further beliefs. Accordingly, I suggest we give up the idea that meaning or knowledge is grounded on something that counts as an ultimate source of evidence. No doubt meaning and knowledge depend on experience, and experience ultimately on sensation. But this is the 'depend' of causality, not of evidence or justification.

I have now stated my problem as well as I can. The search for an empirical foundation for meaning or knowledge leads to skepticism, while a coherence theory seems at a loss to provide any reason for a believer to believe that his beliefs, if coherent, are true. We are caught between a false answer to the skeptic, and no answer.

The dilemma is not a true one. What is needed to answer the skeptic is to show that someone with a (more or less) coherent set of beliefs has a reason to suppose his beliefs are not mistaken in the main. What we have shown is that it is absurd to look for a justifying ground for the totality of beliefs, something outside this totality which we can use to test or compare with our beliefs. The answer to our problem must then be to find a *reason* for supposing most of our beliefs are true that is not a form of *evidence*.

My argument has two parts. First I urge that a correct understanding of the speech, beliefs, desires, intentions and other propositional attitudes of a person leads to the conclusion that most of a person's beliefs must be true, and so there is a legitimate presumption that any one of them, if it coheres with most of the rest, is true. Then I go on to claim that anyone with thoughts, and so in particular anyone who wonders whether he has any reason to suppose he is generally right about the nature of his environment, must know what a belief is, and how in general beliefs are to be detected and interpreted. These being perfectly

general facts we cannot fail to use when we communicate with others, or when we try to communicate with others, or even when we merely think we are communicating with others, there is a pretty strong sense in which we can be said to know that there is a presumption in favor of the overall truthfulness of anyone's beliefs, including our own. So it is bootless for someone to ask for some *further* reassurance; that can only add to his stock of beliefs. All that is needed is that he recognize that belief is in its nature veridical.

Belief can be seen to be veridical by considering what determines the existence and contents of a belief. Belief, like the other so-called propositional attitudes, is supervenient on facts of various sorts, behavioral, neuro-physiological, biological and physical. The reason for pointing this out is not to encourage definitional or nomological reduction of psychological phenomena to something more basic, and certainly not to suggest epistemological priorities. The point is rather understanding. We gain one kind of insight into the nature of the propositional attitudes when we relate them systematically to one another and to phenomena on other levels. Since the propositional attitudes are deeply interlocked, we cannot learn the nature of one by first winning understanding of another. As interpreters, we work our way into the whole system, depending much on the pattern of interrelationships.

Take for example the interdependence of belief and meaning. What a sentence means depends partly on the external circumstances that cause it to win some degree of conviction; and partly on the relations, grammatical, logical or less, that the sentence has to other sentences held true with varying degrees of conviction. Since these relations are themselves translated directly into beliefs, it is easy to see how meaning depends on belief. Belief, however, depends equally on meaning, for the only access to the fine structure and individuation of beliefs is through the sentences speakers and interpreters of speakers use to express and describe beliefs. If we want to illuminate the nature of meaning and belief, therefore, we need to start with something that assumes neither. Quine's suggestion, which I shall essentially follow, is to take *prompted assent* as basic, the causal relation between assenting to a sentence and the cause of such assent. This is a fair place to start the project of identifying beliefs and meanings, since a speaker's assent to a sentence depends both on what he means by the sentence and on what he believes

about the world. Yet it is possible to know that a speaker assents to a sentence without knowing either what the sentence, as spoken by him, means, or what belief is expressed by it. Equally obvious is the fact that once an interpretation has been given for a sentence assented to, a belief has been attributed. If correct theories of interpretation are not unique (do not lead to uniquely correct interpretations), the same will go for attributions of belief, of course, as tied to acquiescence in particular sentences.

A speaker who wishes his words to be understood cannot systematically deceive his would-be interpreters about when he assents to sentences – that is, holds them true. As a matter of principle, then, meaning, and by its connection with meaning, belief also, are open to public determination. I shall take advantage of this fact in what follows and adopt the stance of a radical interpreter when asking about the nature of belief. What a fully informed interpreter could learn about what a speaker means is all there is to learn; the same goes for what the speaker believes.⁷

The interpreter's problem is that what he is assumed to know – the causes of assents to sentences of a speaker – is, as we have seen, the product of two things he is assumed not to know, meaning and belief. If he knew the meanings he would know the beliefs, and if he knew the beliefs expressed by sentences assented to, he would know the meanings. But how can he learn both at once, since each depends on the other?

The general lines of the solution, like the problem itself, are owed to Quine. I will, however, introduce some changes into Quine's solution, as I have into the statement of the problem. The changes are directly relevant to the issue of epistemological skepticism.

I see the aim of radical interpretation (which is much, but not entirely, like Quine's radical translation) as being to produce a Tarski-style characterization of truth for the speaker's language, and a theory of his beliefs. (The second follows from the first plus the presupposed knowledge of sentences held true.) This adds little to Quine's program of translation, since translation of the speaker's language into one's own plus a theory of truth for one's own language add up to a theory of truth for the speaker. But the shift to the semantic notion of truth from the syntactic notion of translation puts the formal restrictions of a theory of truth in the foreground, and emphasizes one aspect of the close relation between truth and meaning.

The principle of charity plays a crucial role in Quine's method, and an even more crucial role in my variant. In either case, the principle directs the interpreter to translate or interpret so as to read some of his own standards of truth into the pattern of sentences held true by the speaker. The point of the principle is to make the speaker intelligible, since too great deviations from consistency and correctness leave no common ground on which to judge either conformity or difference. From a formal point of view, the principle of charity helps solve the problem of the interaction of meaning and belief by restraining the degrees of freedom allowed belief while determining how to interpret words.

We have no choice, Quine has urged, but to read our own logic into the thoughts of a speaker; Quine says this for the sentential calculus, and I would add the same for first-order quantification theory. This leads directly to the identification of the logical constants, as well as to assigning a logical form to all sentences.

Something like charity operates in the interpretation of those sentences whose causes of assent come and go with time and place: when the interpreter finds a sentence of the speaker the speaker assents to regularly under conditions he recognizes, he takes those conditions to be the truth conditions of the speaker's sentence. This is only roughly right, as we shall see in a moment. Sentences and predicates less directly geared to easily detected goings-on can, in Quine's canon, be interpreted at will, given only the constraints of interconnections with sentences conditioned directly to the world. Here I would extend the principle of charity to favor interpretations that as far as possible preserve truth: I think it makes for mutual understanding, and hence for better interpretation, to interpret what the speaker accepts as true when we can. In this matter, I have less choice than Quine, because I do not see how to draw the line between observation sentences and theoretical sentences at the start. There are several reasons for this, but the one most relevant to the present topic is that this distinction is ultimately based on an epistemological consideration of a sort I have renounced: observation sentences are directly based on something like sensation – patterns of sensory stimulation – and this is an idea I have been urging leads to skepticism. Without the direct tie to sensation or stimulation, the distinction between observation sentences and others can't be drawn on epistemologically significant

grounds. The distinction between sentences whose causes to assent come and go with observable circumstances and those a speaker clings to through change remains however, and offers the possibility of interpreting the words and sentences beyond the logical.

The details are not here to the point. What should be clear is that if the account I have given of how belief and meaning are related and understood by an interpreter, then most of the sentences a speaker holds to be true – especially the ones he holds to most stubbornly, the ones most central to the system of his beliefs – most of these sentences *are* true, at least in the opinion of the interpreter. For the only, and therefore unimpeachable, method available to the interpreter automatically puts the speaker's beliefs in accord with the standards of logic of the interpreter, and hence credits the speaker with plain truths of logic. Needless to say there are degrees of logical and other consistency, and perfect consistency is not to be expected. What needs emphasis is only the methodological necessity for finding consistency enough.

Nor, from the interpreter's point of view, is there any way he can discover the speaker to be largely wrong about the world. For he interprets sentences held true (which is not to be distinguished from attributing beliefs) according to the events and objects in the outside world that cause the sentence to be held true.

What I take to be the important aspect of this approach is apt to be missed because the approach reverses our natural way of thinking of communication derived from situations in which understanding has already been secured. Once understanding has been secured we are able, often, to learn what a person believes quite independently of what caused him to believe it. This may lead us to the crucial, indeed fatal, conclusion that we can in general fix what someone means independently of what he believes and independently of what caused the belief. But if I am right, we can't in general first identify beliefs and meanings and then ask what caused them. The causality plays an indispensable role in determining the content of what we say and believe. This is a fact we can be led to recognize by taking up, as we have, the interpreter's point of view.

It is an artifact of the interpreter's correct interpretation of a person's speech and attitudes that there is a large degree of truth and consistency in the thought and speech of an agent. But this is

truth and consistency by the interpreter's standards. Why couldn't it happen that speaker and interpreter understand one another on the basis of shared but erroneous beliefs? This can, and no doubt often does, happen. But it cannot be the rule. For imagine for a moment an interpreter who is omniscient about the world, and about what does and would cause a speaker to assent to any sentence in his (potentially unlimited) repertoire. The omniscient interpreter, using the same method as the fallible interpreter, finds the fallible speaker largely consistent and correct. By his own standards, of course, but since these are objectively correct, the fallible speaker is seen to be largely correct and consistent by objective standards. We may also, if we want, let the omniscient interpreter turn his attention to the fallible interpreter of the fallible speaker. It turns out that the fallible interpreter can be wrong about some things, but not in general; and so he cannot share universal error with the agent he is interpreting. Once we agree to the general method of interpretation I have sketched, it becomes impossible correctly to hold that anyone could be mostly wrong about how things are.

There is, as I noted above, a key difference between the method of radical interpretation I am now recommending, and Quine's method of radical translation. The difference lies in the nature of the choice of causes that govern interpretation. Quine makes interpretation depend on patterns of sensory stimulation, while I make it depend on the external events and objects the sentence is interpreted as being about. Thus Quine's notion of meaning is tied to sensory criteria, something he thinks that can be treated also as evidence. This leads Quine to give epistemic significance to the distinction between observation sentences and others, since observation sentences are supposed, by their direct conditioning to the senses, to have a kind of extra-linguistic justification. This is the view against which I argued in the first part of my paper, urging that sensory stimulations are indeed part of the causal chain that leads to belief, but cannot, without confusion, be considered to be evidence, or a source of justification, for the stimulated beliefs.

What stands in the way of global skepticism of the senses is, in my view, the fact that we must, in the plainest and methodologically most basic cases, take the objects of a belief to be the causes of that belief. And what we, as interpreters, must take them to be is what they in fact are. Communica-

tion begins where causes converge: your utterance means what mine does if belief in its truth is systematically caused by the same events and objects.⁸

The difficulties in the way of this view are obvious, but I think they can be overcome. The method applies directly, at best, only to occasion sentences – the sentences assent to which is caused systematically by common changes in the world. Further sentences are interpreted by their conditioning to occasion sentences, and the appearance in them of words that appear also in occasion sentences. Among occasion sentences, some will vary in the credence they command not only in the face of environmental change, but also in the face of change of credence awarded related sentences. Criteria can be developed on this basis to distinguish degrees of observationality on internal grounds, without appeal to the concept of a basis for belief outside the circle of beliefs.

Related to these problems, and easier still to grasp, is the problem of error. For even in the simplest cases it is clear that the same cause (a rabbit scampers by) may engender different beliefs in speaker and observer, and so encourage assent to sentences which cannot bear the same interpretation. It is no doubt this fact that made Quine turn from rabbits to patterns of stimulation as the key to interpretation. Just as a matter of statistics, I'm not sure how much better one approach is than the other. Is the relative frequency with which identical patterns of stimulation will touch off assent to "Gavagai" and "Rabbit" greater than the relative frequency with which a rabbit touches off the same two responses in speaker and interpreter? Not an easy question to test in a convincing way. But let the imagined results speak for Quine's method. Then I must say, what I must say in any case, the problem of error cannot be met sentence by sentence, even at the simplest level. The best we can do is cope with error holistically, that is, we interpret so as to make an agent as intelligible as possible, given his actions, his utterances and his place in the world. About some things we will find him wrong, as the necessary cost of finding him elsewhere right. As a rough approximation, finding him right means identifying the causes with the objects of his beliefs, giving special weight to the simplest cases, and countenancing error where it can be best explained.

Suppose I am right that an interpreter must so interpret as to make a speaker or agent largely correct about the world. How does this help the

person himself who wonders what reason he has to think his beliefs are mostly true? How can he learn about the causal relations between the real world and his beliefs that lead the interpreter to interpret him as being on the right track?

The answer is contained in the question. In order to doubt or wonder about the provenance of his beliefs an agent must know what belief is. This brings with it the concept of objective truth, for the notion of a belief is the notion of a state that may or may not jibe with reality. But beliefs are also identified, directly and indirectly, by their causes. What an omniscient interpreter knows a fallible interpreter gets right enough if he understands a speaker, and this is just the complicated causal truth that makes us the believers we are, and fixes the contents of our beliefs. The agent has only to reflect on what a belief is to appreciate that most of his basic beliefs are true, and among his beliefs, those most securely held and that cohere with the main body of his beliefs are the most apt to be true. The question, how do I know my beliefs are generally true? thus answers itself, simply because beliefs are by nature generally true. Rephrased or expanded, the question becomes, how can I tell whether my beliefs, which are by their nature generally true, are generally true?

All beliefs are justified in this sense: they are supported by numerous other beliefs (otherwise they wouldn't be the beliefs they are), and have a presumption in favor of their truth. The presump-

tion increases the larger and more significant the body of beliefs with which a belief coheres, and there being no such thing as an isolated belief, there is no belief without a presumption in its favor. In this respect, interpreter and interpreted differ. From the interpreter's point of view, methodology enforces a general presumption of truth for the body of beliefs as a whole, but the interpreter does not need to presume each particular belief of someone else is true. The general presumption applied to others does not make them globally right, as I have emphasized, but provides the background against which to accuse them of error. But from each person's own vantage point, there must be a graded presumption in favor of each of his own beliefs.

We cannot, alas, draw the picturesque and pleasant conclusion that all true beliefs constitute knowledge. For though all of a believer's beliefs are to some extent justified to him, some may not be justified enough, or in the right way, to constitute knowledge. The general presumption in favor of the truth of belief serves to rescue us from a standard form of skepticism by showing why it is impossible for all our beliefs to be false together. This leaves almost untouched the task of specifying the conditions of knowledge. I have not been concerned with the canons of evidential support (if such there be), but to show that all that counts as evidence or justification for a belief must come from the same totality of belief to which it belongs.

Notes

- 1 See my 'True to the Facts', *The Journal of Philosophy* (1960), pp. 216-34.
- 2 Hilary Putnam, *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul, 1978), p. 125.
- 3 See my 'On the Very Idea of a Conceptual Scheme', in *Proceedings and Addresses of the American Philosophical Association* (1974), pp. 5-20.
- 4 Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979), p. 178.
- 5 W. V. Quine, 'The Nature of Natural Knowledge', in S. Guttenplan (ed.), *Mind and Language* (Clarendon Press, Oxford, 1975), p. 68.
- 6 Many other passages in Quine suggest that Quine hopes to assimilate sensory causes to evidence. In *Word and Object* (Massachusetts: MIT Press, 1960), p. 22, he writes that 'surface irritations . . . exhaust our clues to an external world'. In *Ontological Relativity* (New York: Columbia University Press, 1969), p. 75, we find that 'The stimulation of his sensory receptors

is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world.' On the same page: 'Two cardinal tenets of empiricism remain unassailable. . . . One is that whatever evidence there is for science is sensory evidence. The other . . . is that all inculcation of meanings of words, must rest ultimately on sensory evidence.' In *The Roots of Reference* (Illinois: Open Court Publishing Company, 1974), pp. 37-8, Quine says 'observations' are basic 'both in the support of theory and in the learning of language', and then goes on, 'What are observations? They are visual, auditory, tactual, olfactory. They are sensory, evidently, and thus subjective. . . . Should we say then that the observation is not the sensation. . . .? No. . . .' Quine goes on to abandon talk of observations for talk of observation sentences. But of course observation sentences, unlike observations, cannot play the role of evidence unless we have reason to believe they are true.

- 7 I now think it is essential, in doing radical interpretation, to include the desires of the speaker from the start, so that the springs of action and intention, namely both belief and desire, are related to meaning. But in the present talk it is not necessary to introduce this further factor.
- 8 It is clear that the causal theory of meaning has little in common with the causal theories of reference of

Kripke and Putnam. Those theories look to causal relations between names and objects of which speakers may well be ignorant. The chance of systematic error is thus increased. My causal theory does the reverse by connecting the cause of a belief with its object.



PART IV

Epistemic Justification

Introduction

The selections in this section attempt to answer the question "Under what general conditions is one epistemically justified in believing a proposition?" Two key issues to be examined in answering this question are the relation between justification and truth and that between justification and criticizability.

One can plainly be justified in believing that p even if p is not true, but does being justified in believing that p at least make it objectively probable that p ? If it doesn't, why should we care about justification?

The question of whether having unjustified beliefs entails being criticizable might seem to be easily answered. Yes, if one has an unjustified belief, that belief is irrational, and one can be rightly criticized for having an irrational belief. Yet, there is a basic problem with this response. Broadly speaking, epistemologists often want to use the term "epistemic justification" to pick out a kind of positive value status related to gaining truth and avoiding error. The crucial point, emphasized by Richard Feldman and Earl Conee in their selection, is that deontological statuses – statuses of being obligated, being forbidden – do not exhaust positive value statuses. Having negative deontological status does plausibly correlate with being rightly criticizable. If one doesn't do what one ought to do, one may be criticized for not so acting. But it is often acknowledged in ethics, for example, that it is bad to have cruel instincts. One cannot be fairly criticized for having cruel instincts, something that is not in one's power; rather, such instincts have a negative moral or ethical value. Depending on one's theory of such value, this may be a matter of not being conducive

to the production of states of pleasure, states of desire satisfaction, etc. Similarly, a natural proclivity toward wishful thinking, while not being something criticizable, is of low epistemic value.

Feldman and Conee argue in favor of evidentialism about justification. Whether a subject's doxastic attitude toward a proposition is justified is determined by whether taking that attitude fits the subject's evidence. Evidence here includes experiential as well as doxastic evidence. Feldman and Conee defend this view against reliabilist theories which connect justified belief essentially with objective probability of truth. They claim that important reliabilist intuitions can be captured in their framework by appealing to the notion of well-foundedness, a notion which employs the evidentialist conception of justification together with the notion of a basing relation. S 's doxastic attitude D toward p is well-founded if and only if having D toward p is justified and S has D toward p on the basis of a body of evidence e that meets the following conditions: (i) S has e as evidence; (ii) having D toward p fits e ; and (iii) there is no more inclusive body of evidence e' had by S such that having D toward p does not fit e' . Here (iii) is necessary to ensure that S has no undermining justification.

As do Feldman and Conee, Richard Foley insists there is a kind of positive epistemic status that is not essentially related to objective probability of truth. More generally, he sees a basic division of labor in epistemology. Some sorts of epistemic value depend on luck and some do not. Knowledge, for example, depends on a factor of luck, which explains why we cannot have the assurance Descartes sought, the guarantee that our beliefs about the world are true. Yet there is

Introduction

still a kind of luck-free epistemic value that epistemologists may investigate, a kind that depends only on that over which we have substantial, if not full, control. This is what Foley calls egocentric rationality. If I would believe that p upon deep reflection, then it is egocentrically rational for me to believe that p . Whether I am egocentrically rational in believing what I do is thus an epistemic good over which I have control. If I lack egocentric rationality, I am to be criticized, for egocentric rationality is something even a very hostile epistemic environment cannot strip from me. Why seek egocentric rationality, one might ask, if it provides no guarantee of truth or even of objective probability of truth? Foley's answer seems to be that we believe by our best lights that it is effective to pursue our goal of having accurate and comprehensive beliefs by being egocentrically rational. We are working without nets, as Foley puts it, but this is the only way we can work, and it is a good way to work.

John Pollock's epistemology includes the following elements: (i) the notion that concepts are defined by their justification conditions; (ii) the claim that epistemic rules are rules of prima facie justification or the overriding of such justification; (iii) the view that epistemic rules are rules *by which we reason*; and (iv) the thesis that our acceptance of such rules is essentially implicit and procedural, and that it reaches consciousness only through the persistent reflection of epistemology.

Pollock uses an analogy with bicycle riding to explain the sort of unity he sees in criteriology. A bicycle rider embodies procedural rules that he can occasionally override, and these rules are also mostly implicit and subconscious, and they have normative content as well. What unifies such rules, moreover, is simply that they specify how one rides a bike (correctly and, for the most part, actually).

Unlike Feldman and Conee and Foley, Susan Haack aims to connect epistemic justification essentially with truth-conduciveness. After examining varieties of foundationalist and coherentist theories, she claims that we remain in need of an account of how there could be both logical and causal relations between experience and beliefs. Only a logical relation can ensure the rational or justificatory connection between experience and belief. And only a causal connection can ensure the linking of empirical justification with truth. For empirical worldly fact can enter our cognitive economy only through experience.

In her book, Haack proposes to provide both the logical and causal connection by employing a distinction between belief states (S-beliefs) and the contents of those states (C-beliefs). She begins by giving an evidentialist account of justification: agent A is more/less justified in believing that p depending upon how good A 's evidence is for p . The distinction between S- and C-beliefs is then employed in characterizing A 's evidence. A 's evidence consists of three sorts of items: A 's S-reasons, A 's C-reasons, and A 's experiential C-evidence for believing that p . The S-reasons are themselves S-beliefs sustained ultimately by A 's experiential S-states. The role of experience in sustaining S-beliefs, Haack claims, identifies what was right about experientialist foundationalism. A 's C-reasons for believing that p are the C-beliefs that serve as the contents of A 's S-reasons for believing that p . Coherentists were right to emphasize the non-linear character of C-reasons in justification. No class of C-beliefs is basic in the nexus of C-reasons. Finally, A 's experiential C-evidence consists of true propositions to the effect that A is in a certain state, viz. the state that constitutes A 's experiential S-evidence for believing that p . It is the last element of A 's evidence, Haack believes, that supplies the necessary connection between justification and truth. One might be tempted to doubt this: surely, a proposition to the effect that it's visually as if there is something green before me provides no guaranteed link to truth. It could very well be that my experience is unveridical. How could the mere fact that I have an experience as of something green before me be evidence in favor of there being something green before me? Haack's answer is that the appropriate description of the experience characterizes it in a world-involving way. Thus, a visual experience as of there being something green before me is to be described as the kind of experience a normal subject would be in, in normal circumstances, when looking at a green thing. This would seem to provide the link between justification and truth. That I'm in the kind of experiential state that is normally or typically caused by a green thing does seem to make it objectively more probable that there is a green thing before me.

Thus, we have foundherentism. Foundationalist elements survive in the claim that experiential S-reasons form the causal bedrock, coherentist elements in the claim that the structure of C-reasons do not have a linear structure. The connection with truth, missed by coherentism and by many

forms of foundationalism, is secured through the claim that part of the C-evidence for a belief includes truths describing experiences in terms of their typical external causes. (It is interesting to

compare such foundherentism with the views defended by John McDowell in Part VII and those of Ernest Sosa in Part V.)

Further Reading

Alston, William P., *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
 Audi, Robert, *The Structure of Justification* (New York: Cambridge University Press, 1993).
 Foley, R., *The Theory of Epistemic Rationality* (Cambridge, MA: Harvard University Press, 1987).
 —, *Working Without a Net: A Study of Egocentric Epistemology* (Oxford: Oxford University Press, 1993).
 Haack, Susan, *Evidence and Inquiry* (Oxford: Blackwell Publishers, 1993).
 Lehrer, Keith, *Theory of Knowledge* (Boulder, CO: Westview, 1990).
 Lycan, William, *Judgement and Justification* (Cambridge: Cambridge University Press, 1988).

Moser, Paul K., *Knowledge and Evidence* (Cambridge: Cambridge University Press, 1989).
 Pappas, George S. (ed.), *Justification and Knowledge* (Dordrecht: Kluwer Academic Publishers, 1979).
 Plantinga, Alvin, *Warrant: The Current Debate* (Oxford: Oxford University Press, 1993).
 Pollock, John, *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986).
 Sosa, Ernest, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991), Part II.

Evidentialism

Richard Feldman and Earl Conee

I

We advocate evidentialism in epistemology. What we call evidentialism is the view that the epistemic justification of a belief is determined by the quality of the believer's evidence for the belief. Disbelief and suspension of judgment also can be epistemically justified. The doxastic attitude that a person is justified in having is the one that fits the person's evidence. More precisely:

EJ Doxastic attitude *D* toward proposition *p* is epistemically justified for *S* at *t* if and only if having *D* toward *p* fits the evidence *S* has at *t*.¹

We do not offer EJ as an analysis. Rather it serves to indicate the kind of notion of justification that we take to be characteristically epistemic—a notion that makes justification turn entirely on evidence. Here are three examples that illustrate the application of this notion of justification. First, when a physiologically normal person under ordinary circumstances looks at a plush green lawn that is directly in front of him in broad daylight, believing that there is something green before him is the attitude toward this proposition that fits his evidence. That is why the belief is epistemically justified. Second, suspension of judgment is the fitting attitude for each of us toward the proposition that an even number of ducks exists, since our evidence makes it equally likely that the number is odd

Neither belief nor disbelief is epistemically justified when our evidence is equally balanced. And third, when it comes to the proposition that sugar is sour, our gustatory experience makes disbelief the fitting attitude. Such experiential evidence epistemically justifies disbelief.²

EJ is not intended to be surprising or innovative. We take it to be the view about the nature of epistemic justification with the most initial plausibility. A defense of EJ is now appropriate because several theses about justification that seem to cast doubt on it have been prominent in recent literature on epistemology. Broadly speaking, these theses imply that epistemic justification depends upon the cognitive capacities of people, or upon the cognitive processes or information-gathering practices that led to the attitude. In contrast, EJ asserts that the epistemic justification of an attitude depends only on evidence.

We believe that EJ identifies the basic concept of epistemic justification. We find no adequate grounds for accepting the recently discussed theses about justification that seem to cast doubt on EJ. In the remainder of this paper we defend evidentialism. Our purpose is to show that it continues to be the best view of epistemic justification.

II

In this section we consider two objections to EJ. Each is based on a claim about human limits and a claim about the conditions under which an attitude can be justified. One objection depends on the claim that an attitude can be justified only if it is voluntarily adopted, the other depends on the

Originally published in *Philosophical Studies* 48 (1985), pp. 15–34; reprinted with kind permission from Kluwer Academic Publishers.

claim that an attitude toward a proposition or propositions can be justified for a person only if the ability to have that attitude toward the proposition or those propositions is within normal human limits.

Doxastic voluntarism

EJ says that a doxastic attitude is justified for a person when that attitude fits the person's evidence. It is clear that there are cases in which a certain attitude toward a proposition fits a person's evidence, yet the person has no control over whether he forms that attitude toward that proposition. So some involuntarily adopted attitudes are justified according to EJ. John Heil finds this feature of the evidentialist position questionable. He says that the fact that we "speak of a person's beliefs as being warranted, justified, or rational . . . makes it appear that . . . believing something can, at least sometimes, be under the voluntary control of the believer."³ Hilary Kornblith claims that it seems "unfair" to evaluate beliefs if they "are not subject" to "direct voluntary control."⁴ Both Heil and Kornblith conclude that although beliefs are not under *direct* voluntary control, it is still appropriate to evaluate them because "they are not entirely out of our control either."⁵ "One does have a say in the procedures one undertakes that lead to" the formation of beliefs.⁶

Doxastic attitudes need not be under any sort of voluntary control for them to be suitable for epistemic evaluation. Examples confirm that beliefs may be both involuntary and subject to epistemic evaluation. Suppose that a person spontaneously and involuntarily believes that the lights are on in the room, as a result of the familiar sort of completely convincing perceptual evidence. This belief is clearly justified, whether or not the person cannot voluntarily acquire, lose, or modify the cognitive process that led to the belief. Unjustified beliefs can also be involuntary. A paranoid man might believe without any supporting evidence that he is being spied on. This belief might be a result of an uncontrollable desire to be a recipient of special attention. In such a case the belief is clearly epistemically unjustified even if the belief is involuntary and the person cannot alter the process leading to it.

The contrary view that only voluntary beliefs are justified or unjustified may seem plausible if one confuses the topic of EJ with an assessment of

the *person*.⁷ A person deserves praise or blame for being in a doxastic state only if that state is under the person's control.⁸ The person who involuntarily believes in the presence of overwhelming evidence that the lights are on does not deserve praise for this belief. The belief is nevertheless justified. The person who believes that he is being spied on as a result of an uncontrollable desire does not deserve to be blamed for that belief. But there is a fact about the belief's epistemic merit. It is epistemically defective – it is held in the presence of insufficient evidence and is therefore unjustified.

Doxastic limits

Apart from the questions about doxastic voluntarism, it is sometimes claimed that it is inappropriate to set epistemic standards that are beyond normal human limits. Alvin Goldman recommends that epistemologists seek epistemic principles that can serve as practical guides to belief formation. Such principles, he contends, must take into account the limited cognitive capacities of people. Thus, he is led to deny a principle instructing people to believe all the logical consequences of their beliefs, since they are unable to have the infinite number of beliefs that following such a principle would require.⁹ Goldman's view does not conflict with EJ, since EJ does not instruct anyone to believe anything. It simply states a necessary and sufficient condition for epistemic justification. Nor does Goldman think this view conflicts with EJ, since he makes it clear that the principles he is discussing are guides to action and not principles that apply the traditional concept of epistemic justification.

Although Goldman does not use facts about normal cognitive limits to argue against EJ, such an argument has been suggested by Kornblith and by Paul Thagard. Kornblith cites Goldman's work as an inspiration for his view that "having justified beliefs is simply doing the best one can in the light of the innate endowment one starts from."¹⁰ Thagard contends that rational or justified principles of inference "should not demand of a reasoner inferential performance which exceeds the general psychological abilities of human beings."¹¹ Neither Thagard nor Kornblith argues against EJ, but it is easy to see how such an argument would go: A doxastic attitude toward a proposition is justified for a person only if having that attitude

toward that proposition is within the normal doxastic capabilities of people. Some doxastic attitudes that fit a person's evidence are not within those capabilities. Yet EJ classifies them as justified. Hence, EJ is false.

We see no good reason here to deny EJ. The argument has as a premise the claim that some attitudes beyond normal limits do fit someone's evidence. The fact that we are limited to a finite number of beliefs is used to support this claim. But this fact does not establish the premise. There is no reason to think that an infinite number of beliefs fits any body of evidence that anyone ever has. The evidence that people have under ordinary circumstances never makes it evident, concerning every one of an infinite number of logical consequences of that evidence, that it is a consequence. Thus, believing each consequence will not fit any ordinary evidence. Furthermore, even if there are circumstances in which more beliefs fit a person's evidence than he is able to have, all that follows is that he cannot have at one time all the beliefs that fit. It does not follow that there is any particular fitting belief which is unattainable. Hence, the premise of the argument that says that EJ classifies as justified some normally unattainable beliefs is not established by means of this example. There does not seem to be any sort of plausible evidence that would establish this premise. While some empirical evidence may show that people typically do not form fitting attitudes in certain contexts, or that some fitting attitudes are beyond some individual's abilities, such evidence fails to show that any fitting attitudes are beyond normal limits.¹²

There is a more fundamental objection to this argument against EJ. There is no basis for the premise that what is epistemically justified must be restricted to feasible doxastic alternatives. It can be a worthwhile thing to help people to choose among the epistemic alternatives open to them. But suppose that there were occasions when forming the attitude that best fits a person's evidence was beyond normal cognitive limits. This would still be the attitude *justified* by the person's evidence. If the person had normal abilities, then he would be in the unfortunate position of being unable to do what is justified according to the standard for justification asserted by EJ. This is not a flaw in the account of justification. Some standards are met only by going beyond normal human limits. Standards that some teachers set for an "A" in a course are unattainable for most students. There are standards of artistic excellence

that no one can meet, or at least standards that normal people cannot meet in any available circumstance. Similarly, epistemic justification might have been normally unattainable.

We conclude that neither considerations of doxastic voluntarism nor of doxastic limits provide any good reason to abandon EJ as an account of epistemic justification.

III

EJ sets an epistemic standard for evaluating doxastic conduct. In any case of a standard for conduct, whether it is voluntary or not, it is appropriate to speak of "requirements" or "obligations" that the standard imposes. The person who has overwhelming perceptual evidence for the proposition that the lights are on, epistemically ought to believe that proposition. The paranoid person epistemically ought not to believe that he is being spied upon when he has no evidence supporting this belief. We hold the general view that one epistemically ought to have the doxastic attitudes that fit one's evidence. We think that being epistemically obligatory is equivalent to being epistemically justified.

There are in the literature two other sorts of view about epistemic obligations. What is epistemically obligatory, according to these other views, does not always fit one's evidence. Thus, each of these views of epistemic obligation, when combined with our further thesis that being epistemically obligatory is equivalent to being epistemically justified, yields results incompatible with evidentialism. We shall now consider how these proposals affect EJ.

Justification and the obligation to believe truths

Roderick Chisholm holds that one has an "intellectual requirement" to try one's best to bring it about that, of the propositions one considers, one believes all and only the truths.¹³ This theory of what our epistemic obligations are, in conjunction with our view that the justified attitudes are the ones we have an epistemic obligation to hold, implies the following principle:

CJ Doxastic attitude *D* toward proposition *p* is justified for person *S* at time *t* if and only if *S* considers *p* at *t* and *S*'s having *D* toward *p*

at t would result from S's trying his best to bring it about that S believe p at t iff p is true.

Evaluation of CJ is complicated by an ambiguity in "trying one's best." It might mean "trying in that way which will in fact have the best result." Since the goal is to believe all and only the truths one considers, the best results would be obtained by believing each truth one considers and disbelieving each falsehood one considers. On this interpretation, CJ implies that believing each truth and disbelieving each falsehood one considers is justified whenever believing and disbelieving in these ways would result from something one could try to do.

On this interpretation CJ is plainly false. We are not justified in believing every proposition we consider that happens to be true and which we could believe by trying for the truth. It is possible to believe some unsubstantiated proposition in a reckless endeavor to believe a truth, and happen to be right. This would not be an "epistemically justified belief."¹⁴

It might be contended that trying one's best to believe truths and disbelieve falsehoods really amounts to trying to believe and disbelieve in accordance with one's evidence. We agree that gaining the doxastic attitudes that fit one's evidence is the epistemically best way to use one's evidence in trying to believe all and only the truths one considers. This interpretation of CJ makes it nearly equivalent to EJ. There are two relevant differences. First, CJ implies that one can have justified attitudes only toward propositions one actually considers. EJ does not have this implication. CJ is also unlike EJ in implying that an attitude is justified if it would result from the *trying* to form the attitude that fits one's evidence. The attitude that is justified according to EJ is the one that as a matter of fact does fit one's evidence. This seems more plausible. What would happen if one tried to have a fitting attitude seems irrelevant – one might try but fail to form the fitting attitude.

We conclude that the doxastic attitudes that would result from carrying out the intellectual requirement that Chisholm identifies are not the epistemically justified attitudes.

Justification and epistemically responsible action

Another view about epistemic obligations, proposed by Hilary Kornblith, is that we are obligated

to seek the truth and gather evidence in a responsible way. Kornblith also maintains that the justification of a belief depends on how responsibly one carried out the inquiry that led to the belief.¹⁵ We shall now examine how the considerations leading to this view affect EJ.

Kornblith describes a case of what he regards as "epistemically culpable ignorance." It is an example in which a person's belief seems to fit his evidence, and thus it seems to be justified according to evidentialism. Kornblith contends that the belief is unjustified because it results from epistemically irresponsible behavior. His example concerns a headstrong young physicist who is unable to tolerate criticism. After presenting a paper to his colleagues, the physicist pays no attention to the devastating objection of a senior colleague. The physicist, obsessed with his own success, fails even to hear the objection, which consequently has no impact on his beliefs, Kornblith says that after this, the physicist's belief in his own theory is unjustified. He suggests that evidentialist theories cannot account for this fact.

Crucial details of this example are left unspecified, but in no case does it provide a refutation of evidentialism. If the young physicist is aware of the fact that his senior colleague is making an objection, then this fact is evidence he has against his theory, although it is unclear from just this much detail how decisive it would be. So, believing his theory may no longer be justified for him according to a purely evidentialist view. On the other hand, perhaps he remains entirely ignorant of the fact that a senior colleague is objecting to his theory. He might be "lost in thought" – privately engrossed in proud admiration of the paper he has just given – and fail to understand what is going on in the audience. If this happens, and his evidence supporting his theory is just as it was prior to his presentation of the paper, then believing the theory does remain justified for him (assuming that it was justified previously). There is no reason to doubt EJ in the light of this example. It may be true that the young physicist is an unpleasant fellow, and that he lacks intellectual integrity. This is an evaluation of the character of the physicist. It is supported by the fact that in this case he is not engaged in an impartial quest for the truth. But the physicist's character has nothing to do with the epistemic status of his belief in his theory.

Responsible evidence-gathering obviously has some epistemic significance. One serious epistemological question is that of how to engage in a

thoroughgoing rational pursuit of the truth. Such a pursuit may require gathering evidence in responsible ways. It may also be necessary to be open to new ideas, to think about a variety of important issues, and to consider a variety of opinions about such issues. Perhaps it requires, as Bonjour suggests, that one "reflect critically upon one's belief."¹⁶ But everyone has some justified beliefs, even though virtually no one is fully engaged in a rational pursuit of the truth. EJ has no implication about the actions one must take in a rational pursuit of the truth. It is about the epistemic evaluation of attitudes given the evidence one does have, however one came to possess that evidence.

Examples like that of the headstrong physicist show no defect in the evidentialist view. Justified beliefs can result from epistemically irresponsible actions.

Other sorts of obligation

Having acknowledged at the beginning of this section that justified attitudes are in a sense obligatory, we wish to forestall confusions involving other notions of obligations. It is not the case that there is always a *moral* obligation to believe in accordance with one's evidence. Having a fitting attitude can bring about disastrous personal or social consequences. Vicious beliefs that lead to vicious acts can be epistemically justified. This rules out any moral obligation to have the epistemically justified attitude.¹⁷

It is also false that there is always a *prudential* obligation to have each epistemically justified attitude. John Heil discusses the following example.¹⁸ Sally has fairly good evidence that her husband Burt has been seeing another woman. Their marriage is in a precarious condition. It would be best for Sally if their marriage were preserved. Sally foresees that, were she to believe that Burt has been seeing another woman, her resulting behavior would lead to their divorce. Given these assumptions, EJ counts as justified at least some measure of belief by Sally in the proposition that Burt has been seeing another woman. But Sally would be better off if she did not have this belief, in light of the fact that she would be best served by their continued marriage. Heil raises the question of what Sally's prudential duty is in this case. Sally's *epistemic* obligation is to believe that her husband is unfaithful. But that gives no reason to deny what seems obvious here. Sally *prudentially* ought to

refrain from believing her husband to be unfaithful. It can be prudent not to have a doxastic attitude that is correctly said by EJ to be justified, just as it can be moral not to have such an attitude.

More generally, the causal consequences of having an unjustified attitude can be more beneficial in *any* sort of way than the consequences of having its justified alternative. We have seen that it can be morally and prudentially best not to have attitudes justified according to EJ. Failing to have these attitudes can also have the best results for the sake of *epistemic* goals such as the acquisition of knowledge. Roderick Firth points out that a scientist's believing against his evidence that he will recover from an illness may help to effect a recovery and so contribute to the growth of knowledge by enabling the scientist to continue his research.¹⁹ William James's case for exercising "the will to believe" suggests that some evidence concerning the existence of God is available only after one believes in God in the absence of justifying evidence. EJ does not counsel against adopting such beliefs for the sake of these epistemic ends. EJ implies that the beliefs would be unjustified when adopted. This is not to say that the believing would do no epistemic good.

We acknowledge that it is appropriate to speak of epistemic obligations. But it is a mistake to think that what is epistemically obligatory, i.e., epistemically justified, is also morally or prudentially obligatory, or that it has the overall best epistemic consequences.

IV

Another argument that is intended to refute the evidentialist approach to justification concerns the ways in which a person can come to have an attitude that fits his evidence. Both Kornblith and Goldman propose examples designed to show that merely *having* good evidence for a proposition is not sufficient to make believing that proposition justified.²⁰ We shall work from Kornblith's formulation of the argument, since it is more detailed. Suppose Alfred is justified in believing *p*, and justified in believing if *p* then *q*. Alfred also believes *q*. EJ seems to imply that believing *q* is justified for Alfred, since that belief does seem to fit this evidence. Kornblith argues that Alfred's belief in *q* may still not be justified. It is not justified, according to Kornblith, if Alfred has a strong distrust of *modus ponens* and believes *q*

because he likes the sound of the sentence expressing it rather than on the basis of the *modus ponens* argument. Similarly, Goldman says that a person's belief in q is not justified unless the belief is caused in some appropriate way.

Whether EJ implies that Alfred's belief in q is justified depends in part on an unspecified detail – Alfred's evidence concerning *modus ponens*. It is possible that Alfred has evidence against *modus ponens*. Perhaps he has just seen a version of the Liar paradox that seems to render *modus ponens* as suspect as the other rules and premises in the derivation. In the unlikely event that Alfred has such evidence, EJ implies that believing q is *not* justified for him. If rather, as we shall assume, his overall evidence supports *modus ponens* and q , then EJ does imply that believing q is justified for him.

When Alfred has strong evidence for q , his believing q is epistemically justified. This is the sense of "justified" captured by EJ. However, if Alfred's basis for believing q is not his evidence for it, but rather the sound of the sentence expressing q , then it seems equally clear that there is some sense in which this state of believing is epistemically "defective" – he did not arrive at the belief in the right way. The term "well-founded" is sometimes used to characterize an attitude that is epistemically both well-supported and properly arrived at. Well-foundedness is a second evidentialist notion used to evaluate doxastic states. It is an evidentialist notion because its application depends on two matters of evidence – the evidence one *has*, and the evidence one *uses* in forming the attitude. More precisely:

- WF S's doxastic attitude D at t toward proposition p is well-founded if and only if
- (i) having D toward p is justified for S at t ; and
 - (ii) S has D toward p on the basis of some body of evidence e , such that
 - (a) S has e as evidence at t ;
 - (b) having D toward p fits e ; and
 - (c) there is no more inclusive body of evidence e' had by S at t such that having D toward p does not fit e' .²¹

Since the evidentialist can appeal to this notion of well-foundedness, cases in which a person has but does not use justifying evidence do not refute evidentialism. Kornblith and Goldman's intuitions about such cases can be accommodated. A person in Alfred's position *is* in an epistemically defective

state – his belief in q is not well-founded. Having said this, it is reasonable also to affirm the other evidentialist judgment that Alfred's belief in q is in another sense epistemically right – it is justified.²²

V

The theory of epistemic justification that has received the most attention recently is reliabilism. Roughly speaking, this is the view that epistemically justified beliefs are the ones that result from belief-forming processes that reliably lead to true beliefs.²³ In this section we consider whether reliabilism casts doubt on evidentialism.

Although reliabilists generally formulate their view as an account of epistemic justification, it is clear that in its simplest forms it is better regarded as an account of well-foundedness. In order for a belief to be favorably evaluated by the simple sort of reliabilism sketched above, the belief must actually be held, as is the case with WF. And just as with WF, the belief must be "grounded" in the proper way. Where reliabilism appears to differ from WF is over the conditions under which a belief is properly grounded. According to WF, this occurs when the belief is based on fitting evidence. According to reliabilism, a belief is properly grounded if it results from a belief-forming process that reliably leads to true beliefs. These certainly are *conceptually* different accounts of the grounds of well-founded beliefs.

In spite of this conceptual difference, reliabilism and WF may be extensionally equivalent. The question of equivalence depends on the resolution of two unclaritys in reliabilism. One pertains to the notion of a belief-forming process and the other to the notion of reliability.

An unclarity about belief-forming processes arises because every belief is caused by a sequence of particular events which is an instance of many types of causal processes. Suppose that one evening Jones looks out of his window and sees a bright shining disk-shaped object. The object is in fact a luminous frisbee, and Jones clearly remembers having given one of these to his daughter. But Jones is attracted to the idea that extraterrestrials are visiting the Earth. He manages to believe that he is seeing a flying saucer. Is the process that caused this belief reliable? Since the sequence of events leading to his belief is an instance of many types of process, the answer depends upon which of these many types is the

relevant one. The sequence falls into highly general categories such as perceptually-based belief formation and visually-based belief formation. It seems that if these are the relevant categories, then his belief is indeed reliably formed, since these are naturally regarded as “generally reliable” sorts of belief-forming processes. The sequence of events leading to Jones’s belief also falls into many relatively specific categories such as night-vision-of-a-nearby-object and vision-in-Jones’s-precise-environmental-circumstances. These are not clearly reliable types. The sequence is also an instance of this contrived kind: process-leading-from-obviously-defeated-evidence-to-the-belief-that-one-sees-a-flying-saucer. This, presumably, is an unreliable kind of process. Finally, there is the maximally specific process that occurs only when physiological events occur that are exactly like those that led to Jones’s belief that he saw a flying saucer. In all likelihood this kind of process occurred only once. Processes of these types are of differing degrees of reliability, no matter how reliability is determined. The implications of reliabilism for the case are rendered definite only when the kind of process whose reliability is relevant is specified. Reliabilists have given little attention to this matter, and those that have specified relevant kinds have not done so in a way that gives their theory in intuitively acceptable extension.²⁴

The second unclarity in reliabilism concerns the notion of reliability itself. Reliability is fundamentally a property of kinds of belief-forming processes, not of sequences of particular events. But we can say that a sequence is reliable provided its relevant type is reliable. The problem raised above concerns the specification of relevant types. The current problem is that of specifying the conditions under which a kind of process is *reliable*. Among possible accounts is one according to which a kind of process is reliable provided most instances of that kind until now have led to true beliefs. Alternative accounts measure the reliability of a kind of process by the frequency with which instances of it produce true beliefs in the future as well as the past, or by the frequency with which its instances produce true beliefs in possible worlds that are similar to the world of evaluation in some designated respect, or by the frequency with which its instances produce true beliefs in all possible worlds.²⁵

Because there are such drastically different ways of filling in the details of reliabilism the application of the theory is far from clear. The possible ver-

sions of reliabilism seem to include one that is extensionally equivalent to WF. It might be held that all beliefs are formed by one of two relevant kinds of belief-forming process. One kind has as instances all and only those sequences of events leading to a belief that is based on fitting evidence; the other is a kind of process that has as instances all and only those sequences leading to a belief that is not based on fitting evidence. If a notion of reliability can be found on which the former sort of process is reliable and the latter is not, the resulting version of reliabilism would be very nearly equivalent to WF.²⁶ We do not claim that reliabilists would favor this version of reliabilism. Rather, our point is that the fact that this *is* a version shows that reliabilism may not even be a rival to WF.²⁷

Evaluation of reliabilism is further complicated by the fact that reliabilists seem to differ about whether they *want* their theory to have approximately the same extension as WF in fact has. The credibility of reliabilism and its relevance to WF depend in part on the concept reliabilists are really attempting to analyze. An example first described by Laurence Bonjour helps to bring out two alternatives.²⁸ Bonjour’s example is of a person who is clairvoyant. As a result of his clairvoyance he comes to believe that the President is in New York City. The person has no evidence showing that he is clairvoyant and no other evidence supporting his belief about the President. Bonjour claims that the example is a counter-example to reliabilism, since the clairvoyant’s belief is not justified (we would add: and therefore ill-founded), although the process that caused it is reliable – the person really is clairvoyant.

The general sort of response to this example that seems to be most commonly adopted by reliabilists is in effect to agree that such beliefs are not well-founded. They interpret or revise reliabilism with the aim of avoiding the counter-example.²⁹ An alternative response would be to argue that the reliability of clairvoyance shows that the belief *is* well-founded, and thus that the example does not refute reliabilism.³⁰

We are tempted to respond to the second alternative – beliefs such as that of the clairvoyant in Bonjour’s example really are well-founded – that this is so clear an instance of an ill-founded belief that any proponent of that view must have in mind a different concept from the one we are discussing. The clairvoyant has no reason for holding his belief about the President. The fact that the belief

was caused by a process of a reliable kind – clairvoyance – is a significant fact about it. Such a belief may merit some favorable term of epistemic appraisal, e.g., “objectively probable.” But the belief is not well-founded.

There are, however, two lines of reasoning that could lead philosophers to think that we must reconcile ourselves to the clairvoyant’s belief turning out to be well-founded. According to one of these arguments, examples such as that of Alfred (discussed in Section IV above) show that the evidentialist account of epistemic merit is unsatisfactory and that epistemic merit must be understood in terms of the reliability of belief-forming processes.³¹ Since the clairvoyant’s belief is reliably formed, our initial inclination to regard it as ill-founded must be mistaken.

This argument is unsound. The most that the example about Alfred shows is that there is a concept of favorable epistemic appraisal other than justification, and that this other concept involves the notion of the *basis* of a belief. We believe that WF satisfactorily captures this other concept. There is no need to move to a reliabilist account, according to which some sort of causal reliability is *sufficient* for epistemic justification. The Alfred example does not establish that some version of reliabilism is correct. It does not establish that the clairvoyant’s belief is well-founded.

The second argument for the conclusion that the clairvoyant’s belief is well-founded makes use of the strong similarity between clairvoyance in Bonjour’s example and normal perception. We claim that Bonjour’s clairvoyant is not justified in his belief about the President because that belief does not fit his evidence. Simply having a spontaneous unferred belief about the whereabouts of the President does not provide evidence for its truth. But, it might be asked, what better evidence is there for any ordinary perceptual belief, say, that one sees a book? If there is no relevant epistemological difference between ordinary perceptual beliefs and the clairvoyant’s belief, then they should be evaluated similarly. The argument continues with the point that reliabilism provides an explanation of the crucial similarity between ordinary perceptual beliefs and the clairvoyant’s belief – both perception and clairvoyance *work*, in the sense that both are reliable. So beliefs caused by each process are well-founded on a reliabilist account. The fact that reliabilism satisfactorily explains this is to the theory’s credit. On the other hand, in advocating evidentialism we have

claimed that perceptual beliefs are well-founded and that the clairvoyant’s belief is not. But there appears to be no relevant evidential difference between these beliefs. Thus, if the evidentialist view of the matter cannot be defended, then reliabilism is the superior theory and we should accept its consequence – the clairvoyant’s belief is well-founded.

One problem with this argument is that reliabilism has no satisfactory explanation of *anything* until the unclarity discussed above are removed in an acceptable way: What shows that perception and clairvoyance are relevant and reliable types of processes? In any event, there *is* an adequate evidentialist explanation of the difference between ordinary perceptual beliefs and the clairvoyant’s belief. On one interpretation of clairvoyance, it is a process whereby one is caused to have beliefs about objects hidden from ordinary view without any conscious state having a role in the causal process. The clairvoyant does not have the conscious experience of, say, seeming to see the President in some characteristic New York City setting, and on that basis form the belief that he is in New York. In this respect, the current version of clairvoyance is unlike ordinary perception, which does include conscious perceptual states. Because of this difference, ordinary perceptual beliefs are based on evidence – the evidence of these sensory states – whereas the clairvoyant beliefs are not based on evidence. Since WF requires that well-founded beliefs be based on fitting evidence, and typical clairvoyant beliefs on the current interpretation are not based on any evidence at all, the clairvoyant beliefs do not satisfy WF.

Suppose instead that clairvoyance does include visual experiences, though of remote objects that cannot stimulate the visual system in any normal way. Even if there are such visual experiences that could serve as a basis for a clairvoyant’s beliefs, still there is a relevant epistemological difference between beliefs based on normal perceptual experience and the clairvoyant’s belief in Bonjour’s example. We have collateral evidence to the effect that when we have perceptual experience of certain kinds, external conditions of the corresponding kinds normally obtain. For example, we have evidence supporting the proposition that when we have the usual sort of experience of seeming to see a book, we usually do in fact see a book. This includes evidence from the coherence of these beliefs with beliefs arising from other

perceptual sources, and it also includes testimonial evidence. This latter point is easily overlooked. One reason that the belief that one sees a book fits even a child's evidence when she has a perceptual experience of seeing a book is that children are taught, when they have the normal sort of visual experiences, that they are seeing a physical object of the relevant kind. This testimony, typically from people whom the child has reason to trust, provides evidence for the child. And of course testimony from others during adult life also gives evidence for the veridicality of normal visual experience. On the other hand, as BonJour describes his example, the clairvoyant has no confirmation at all of his clairvoyant beliefs. Indeed, he has evidence against these beliefs, since the clairvoyant perceptual experiences do not cohere with his other experiences. We conclude, therefore, that evidentialists can satisfactorily explain why ordinary perceptual beliefs are typically well-founded and unconfirmed clairvoyant beliefs, even if reliably caused, are not. There is no good reason to abandon our initial intuition that the beliefs such as those of the clairvoyant in BonJour's example are not well-founded.

Again, reliabilists could respond to BonJour's example either by claiming that the clairvoyant's belief is in fact well-founded or by arguing that reliabilism does not imply that it is well-founded. We turn now to the second of these alternatives, the one most commonly adopted by reliabilists. This view can be defended by arguing either that reliabilism can be reformulated so that it lacks this implication, or that as currently formulated it lacks this implication. We pointed out above that as a general approach reliabilism is sufficiently indefinite to allow interpretations under which it does lack the implication in question. The only way to achieve this result that we know of that is otherwise satisfactory requires the introduction of evidentialist concepts. The technique is to specify the relevant types of belief-forming processes in evidentialist terms. It is possible to hold that the relevant types of belief-forming process are believing something on the basis of fitting evidence and believing not as a result of fitting evidence. This sort of "reliabilism" is a roundabout approximation of the straightforward evidentialist thesis, WF. We see no reason to couch the approximated evidentialist theory in reliabilist terms. Moreover, the reliabilist approximation is not exactly equivalent to WF, and where it differs it appears to go wrong. The difference is this: it seems possible for

the process of believing on the basis of fitting evidence to be unreliable. Finding a suitable sort of reliability makes all the difference here. In various possible worlds where our evidence is mostly misleading, the frequency with which fitting evidence causes true belief is low. Thus, this type of belief-forming process is not "reliable" in such worlds in any straightforward way that depends on actual frequencies. Perhaps a notion of reliability that avoids this result can be found. We know of no such notion which does not create trouble elsewhere for the theory. So, the reliabilist view under consideration has the consequence that in such worlds beliefs based on fitting evidence are not well-founded. This is counterintuitive.³²

In this section we have compared reliabilism and evidentialism. The vagueness of reliabilism makes it difficult to determine what implications the theory has and it is not entirely clear what implications reliabilists want their theory to have. If reliabilists want their theory to have approximately the same extension as WF, we see no better way to accomplish this than one which makes the theory an unnecessarily complex and relatively implausible approximation to evidentialism. If, on the other hand, reliabilists want their theory to have an extension which is substantially different from that of WF, and yet some familiar notion of "a reliable kind of process" is to be decisive for their notion of well-foundedness, then it becomes clear that the concept they are attempting to analyze is not one evidentialists seek to characterize. This follows from the fact that on this alternative they count as well-founded attitudes that plainly do not exemplify the concept evidentialists are discussing. In neither case, then, does reliabilism pose a threat to evidentialism.

VI

Summary and conclusion

We have defended evidentialism. Some opposition to evidentialism rests on the view that a doxastic attitude can be justified for a person only if forming the attitude is an action under the person's voluntary control. EJ is incompatible with the conjunction of this sort of doxastic voluntarism and the plain fact that some doxastic states that fit a person's evidence are out of that person's control. We have argued that no good reason has been given for thinking that an attitude is episte-

mically justified only if having it is under voluntary control.

A second thesis contrary to EJ is that a doxastic attitude can be justified only if having that attitude is within the normal doxastic limits of humans. We have held that the attitudes that are epistemically justified according to EJ are within these limits, and that even if they were not, that fact would not suffice to refute EJ.

Some philosophers have contended that believing a proposition, p , is justified for S only when S has gone about gathering evidence about p in a responsible way, or has come to believe p as a result of seeking a meritorious epistemic goal such as the discovery of truth. This thesis conflicts with EJ, since believing p may fit one's evidence no matter how irresponsible one may have been in

seeking evidence about p and no matter what were the goals that led to the belief. We agree that there is some epistemic merit in responsibly gathering evidence and in seeking the truth. But we see no reason to think that epistemic justification turns on such matters.

Another thesis conflicting with EJ is that merely having evidence is not sufficient to justify belief, since the believer might not make proper use of the evidence in forming the belief. Consideration of this claim led us to make use of a second evidentialist notion, well-foundedness. It does not, however, provide any good reason to think that EJ is false. Nor do we find reason to abandon evidentialism in favor of reliabilism. Evidentialism remains the most plausible view of epistemic justification.

Notes

- 1 EJ is compatible with the existence of varying strengths of belief and disbelief. If there is such variation, then the greater the preponderance of evidence, the stronger the doxastic attitude that fits the evidence.
- 2 There are difficult questions about the concept of fit, as well as about what it is for someone to *have* something as evidence, and of what kind of thing constitutes evidence. As a result, there are some cases in which it is difficult to apply EJ. For example, it is unclear whether a person has as evidence propositions he is not currently thinking of, but could recall with some prompting. As to what constitutes evidence, it seems clear that this includes both beliefs and sensory states such as feeling very warm and having the visual experience of seeing blue. Some philosophers seem to think that only beliefs can justify beliefs. (See, for example, Keith Lehrer, *Knowledge* (Oxford: Oxford University Press, 1974), pp. 187–8.) The application of EJ is clear enough to do the work that we intend here – a defense of the evidentialist position.
- 3 See “Doxastic agency,” *Philosophical Studies* 43 (1983), pp. 355–64. The quotation is from p. 355.
- 4 See “The psychological turn,” *Australasian Journal of Philosophy* 60 (1982), pp. 238–53. The quotation is from p. 252.
- 5 *Ibid.*, p. 253.
- 6 Heil, “Doxastic agency,” p. 363.
- 7 Kornblith may be guilty of this confusion. He writes, “if a person has an unjustified belief, that person is epistemically culpable,” “The psychological turn,” p. 243.
- 8 Nothing we say here should be taken to imply that any doxastic states are in fact voluntarily entered.
- 9 See “Epistemics: The regulative theory of cognition,” *The Journal of Philosophy* LXXV (1978), pp. 509–23, esp. p. 510 and p. 514.
- 10 “Justified belief and epistemically responsible action,” *The Philosophical Review* 92 (1983), pp. 33–48. The quotation is from p. 46.
- 11 Paul Thagard, “From the descriptive to the normative in psychology and logic,” *Philosophy of Science* 49 (1982), pp. 24–42. The quotation is from p. 34.
- 12 Another version of this argument is that EJ is false because it classifies as justified for a person attitudes that are beyond *that person's* limits. This version is subject to similar criticisms.
- 13 See *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977), especially pp. 12–15.
- 14 Roderick Firth makes a similar point against a similar view in “Are epistemic concepts reducible to ethical concepts,” in Alvin Goldman and J. Kim *Values and Morals* (D. Reidel: Dordrecht, 1978), pp. 215–29.
- 15 Kornblith defends this view in “Justified belief and epistemically responsible action.” Some passages suggest that he intends to introduce a new notion of justification, one to be understood in terms of epistemically responsible action. But some passages, especially in Section II, suggest that the traditional analysis of justification is being found to be objectionable and inferior to the one he proposes.
- 16 Laurence Bonjour, “Externalist theories of empirical justification,” *Midwest Studies* V (1980), p. 63.
- 17 This is contrary to the view of Richard Gale, defended in “William James and the ethics of belief,” *American Philosophical Quarterly* 17 (1980), pp. 1–14, and of W. K. Clifford who said, “It is

- wrong always, everywhere, and for every one, to believe anything upon insufficient evidence" (quoted by William James in "The will to believe," reprinted in J. Feinberg (ed.), *Reason and Responsibility*, (Belmont, California: Wadsworth Publishing Co., 1981) p. 100).
- 18 See "Believing what one ought," *Journal of Philosophy* 80, pp. 752ff.
 - 19 See "Epistemic merit, intrinsic and instrumental," *Proceedings and Addresses of The American Philosophical Association* 55 (1981), pp. 5–6.
 - 20 See Kornblith's "Beyond foundationalism and the coherence theory," *The Journal of Philosophy* 77 (1980), pp. 597–612, esp. pp. 601 ff. and Goldman's "What is justified belief?" this vol., ch. 27.
 - 21 Clause (ii) of WF is intended to accommodate the fact that a well-founded attitude need not be based on a person's whole body of evidence. What seems required is that the person base a well-founded attitude on a justifying part of the person's evidence, and that he not ignore any evidence he has that defeats the justifying power of the evidence he does base his attitude on. It might be that his defeating evidence is itself defeated by a still wider body of his evidence. In such a case, the person's attitude is well-founded only if he takes the wider body into account.
WF uses our last main primitive concept – that of basing an attitude on a body of evidence. This notion is reasonably clear, though an analysis would be useful. See Note 22 below for one difficult question about what is entailed.
 - 22 Goldman uses this sort of example only to show that there is a causal element in the concept of justification. We acknowledge that there is an epistemic concept – well-foundedness – that appeals to the notion of basing an attitude on evidence, and this may be a causal notion. What seems to confer epistemic merit on basing one's belief on the evidence is that in doing one appreciates the evidence. It is unclear whether one can appreciate the evidence without being caused to have the belief by the evidence. But in any event we see no such causal requirement in the case of justification.
 - 23 The clearest and most influential discussion of reliabilism is in Goldman's "What is justified belief?" One of the first statements of the theory appears in David Armstrong's *Belief, Truth and Knowledge* (London: Cambridge University Press, 1973). For extensive bibliographies on reliabilism, see Frederick Schmitt's "Reliability, objectivity, and the background of justification," *Australasian Journal of Philosophy* 62 (1984), pp. 1–15, and Richard Feldman's "Reliability and justification," *The Monist* 68 (1985), pp. 159–74.
 - 24 For discussion of the problem of determining relevant kinds of belief-forming processes, see Goldman, "What is justified belief?," Schmitt, "Reliability, objectivity, and the background of justification," Feldman, "Reliability and justification," and Feldman, "Schmitt on reliability, objectivity, and justification," *Australasian Journal of Philosophy* 63 (1985), pp. 354–60.
 - 25 In "Reliability and justified belief," *Canadian Journal of Philosophy* 14, (1984), pp. 103–15, John Pollock argues that there is no account of reliability suitable for reliabilists.
 - 26 This version of reliabilism will not be exactly equivalent to WF because it ignores the factors introduced by clause (ii) of WF.
 - 27 It is also possible that versions of reliabilism making use only of natural psychological kinds of belief-forming processes are extensionally equivalent to WF. Goldman seeks to avoid evaluative epistemic concepts in his theory of epistemic justification, so he would not find an account of justification satisfactory unless it appealed only to such natural kinds. See "What is justified belief?"
 - 28 See "Externalist theories of empirical justification," p. 62.
 - 29 See Goldman, "What is justified belief?," this vol., pp. 362–4, Kornblith, "Beyond foundationalism and the coherence theory," pp. 609–11, and Frederick Schmitt, "Reliability, objectivity, and the background of justification."
 - 30 We know of one who has explicitly taken this approach. It seems to fit most closely with the view defended by David Armstrong in *Belief, Truth and Knowledge*.
 - 31 We know of no one who explicitly defends this inference. In "The psychological turn," p. 241 ff., Kornblith argues that these examples show that justification depends upon "psychological connections" and "the workings of the appropriate belief forming process." But he clearly denies there that reliabilism is directly implied.
 - 32 Stewart Cohen has made this point in "Justification and truth," *Philosophical Studies* 46 (1984), pp. 279–95. Cohen makes the point in the course of developing a dilemma. He argues that reliabilism has the sort of flaw that we describe above when we appeal to worlds where evidence is mostly misleading. Cohen also contends that reliabilism has the virtue of providing a clear explanation of how the epistemic notion of justification is connected with the notion of truth. A theory that renders this truth connection inexplicable is caught on the second horn of Cohen's dilemma.
Although Cohen does not take up evidentialism as we characterize it, the second horn of his dilemma affects EJ and WF. They do not explain how having an epistemically justified or well-founded belief is connected to the truth of that belief. Evidentialists can safely say this much about the truth connection: evidence that makes believing *p* justified is evidence on which it is epistemically probable that *p* is true.

Although there is this connection between justification and truth, we acknowledge that there may be no analysis of epistemic probability that makes the connection to truth as close, or as clear, as might have been hoped.

Cohen argues that there must be a truth connection. This shows no flaw in EJ or WF unless they are

incompatible with there being such a connection. Cohen does not argue for this incompatibility and we know of no reason to believe that it exists. So at most Cohen's dilemma shows that evidentialists have work left to do.

Skepticism and Rationality

Richard Foley

Skeptical hypotheses have been allowed to set the terms of the epistemological debate. They convince no one. Yet they have an enormous influence. It is often influence by provocation. They provoke epistemologists into endorsing metaphysical and linguistic positions that antecedently would have seemed to have had little appeal. Skeptical hypotheses, it is said, cannot even be meaningfully asserted, or if they can, the nature of God or the nature of objects or the nature of thought makes it altogether impossible for them to be true. There are those who refuse to be provoked, but even their epistemologies tend to be dominated by skeptical hypotheses. The hypotheses push them into an overly defensive posture from which it can seem that the test of an epistemology is how well it would fare in a hostile environment. There must be a third way. There must be a way to think about skeptical hypotheses that is neither dismissive nor submissive.

The kind of skeptical challenge that is most familiar to us is the kind that concerned Descartes. To be sure, the skeptical tradition is an ancient one, but the challenges of the ancient skeptics had a different aim from those discussed by Descartes. The followers of Pyrrho of Elis, for example, saw skepticism as a way of life and a desirable one at that. Suspending judgement about how things really are was thought to be a means to tranquillity. There is no hint of this in Descartes or in the Enlightenment philosophers who succeeded him. Descartes did think that skeptical doubt could be

put to good use. It could help deliver us from prejudices and thereby help put our beliefs upon a secure foundation. But even for Descartes, skepticism was first and foremost a threat rather than an opportunity, and it remains so for us. However, Descartes thought that it was a threat that could be successfully met. He thought that by making rational use of our cognitive resources, we can be guaranteed of the truth. Correspondingly, he thought that error is something for which we are always responsible. We have the tools to avoid it. Knowledge is ours for the taking. We need only to be sufficiently reflective and sufficiently cautious. For if we are sufficiently reflective we will come to perceive clearly and distinctly the truth of various claims, and if we are sufficiently cautious we will refrain from believing anything else. Skeptical hypotheses were of interest for Descartes because they provided him with a dramatic way to illustrate these assumptions. They helped him to dramatize the potential power of reason. One need not rely upon tradition or authority for one's opinions. One can stand alone intellectually, deciding for oneself what to make of the world and what to make of one's tradition. And if in doing so one makes proper use of one's reason, one can be assured of knowledge.

An increasing specialization of intellectual labor has made us sensitive, in a way in which Descartes was not, about the extent to which we rely upon the opinions of others, just as a heightened appreciation of cultural relativity has made us more sensitive about the extent to which we are shaped by our traditions. Even so, we are as reluctant to rely uncritically upon our authorities and traditions as Descartes and his Enlightenment succes-

Originally published in M. D. Roth and G. Ross (eds), *Doubting* (Dordrecht: Kluwer Academic Publishers, 1990), pp. 69–81; reprinted with kind permission from Kluwer Academic Publishers.

sors were upon theirs. We realize how difficult it is to distance ourselves intellectually from our surroundings, but we realize also that even our best scientists can be mistaken and that even our most venerable traditions can be misguided. As a result, we too feel the need to make up our own minds. This creates for us an intellectual predicament that is much like the one that Descartes describes at the beginning of the *Meditations*. It is an egocentric predicament, prompted by a simple question, "What am I to believe?". I cannot simply read off from the world what is true, nor can I unproblematically rely upon the acknowledged experts or the received traditions to guide me towards the truth. I instead must marshal my own resources. I must marshal them to determine by my own lights what is true and who is reliable and what if anything is defensible about the traditions of my community. In this respect the individualism of Descartes has won the day.

What we find unacceptable in Descartes is his optimism. We think it naive. We no longer think that by properly marshalling our resources we can be assured of the truth. Being sufficiently reflective and sufficiently cautious is no guarantee that we will avoid error. It is not even a guarantee of reliability. Even so, philosophical problems come down to us through time, and today we remain under the spell of the epistemological aspirations of Descartes, Locke, Hume, Kant and others. The cure is to remind ourselves that their aims need not be ours. What they took to be an intellectual problem in need of a solution we can appreciate as part of the human condition. Given the kind of creatures that we are, we cannot help but lack guarantees of the sort that they sought. This is no more a problem for us than is that of finding a way to do without oxygen. We just are creatures who need oxygen. Similarly, the lack of intellectual guarantees just is part of the human condition. The problem is one of how to react to that condition.

The reaction need not be one of abandoning egocentric epistemology. Reliabilism, for example, constitutes such an abandonment. The egocentric question is "What am I to believe?". To answer this question, I must marshal my resources in an effort to determine what methods of inquiry are reliable. So, from the egocentric perspective, it is altogether unhelpful to be told that I am to have beliefs that are the products of reliable methods. Of course, no sensible reliabilist would claim otherwise. The point, rather, is that reliabilists tend to be satisfied with an epistemology that

does not address the problems of the egocentric predicament, despite the fact that such problems have been at the heart of the great epistemological projects of the past. My point, in turn, is that we need not be satisfied with such an epistemology. We can do better.

But if we are to do better, we must give up an assumption that has had a hold on epistemologists from Descartes through Gettier. According to Descartes, it is rational to believe just that which is clear and distinct for you, and what is clear and distinct for you is true. So, for Descartes rational belief always results in knowledge. This means that there are no Gettier problems within Cartesian epistemology. No one can rationally infer a truth from a rational but false belief, since there are no rational false beliefs. Today we standardly construe the link between rational belief and true belief in a looser manner than did Descartes. A rational belief can be false. So, Gettier problems do arise within our epistemologies. Even so, the difference between Cartesian and contemporary epistemologies is not so great, since within the latter it still is commonly assumed that a rational true belief absent Gettier problems is always knowledge. It is this assumption that must be abandoned. More exactly, it must be abandoned if the answer to the question "What is it rational for me to believe?" is to be relevant to the egocentric predicament. The assumption must be abandoned because it ties rational belief too closely with knowledge and, as a consequence, too closely with reliability. For if by being rational one cannot be assured of having mostly true beliefs, then, contrary to the assumption, a rational true belief need not be a good candidate for knowledge even absent Gettier problems.

Skeptical hypotheses can help illustrate this. Imagine a world in which a demon alters your environment so that you make massive errors about it. You regularly make perceptual mistakes. Even so, the demon allows you to have a few isolated true beliefs about your environment. Perhaps the demon permits the existence of only one chair and it is the one that you are now sitting upon. So, your belief that you are now sitting upon a chair is true. Yet almost all of your other beliefs about your environment are false. This true belief of yours is not a particularly good candidate for knowledge, but why not? There need not be Gettier problems here. You need not have inferred the truth that you are now sitting upon a chair from any falsehood. But then, on the assumption that

rational true belief absent Gettier problems is knowledge, the explanation must be that your belief is not rational. But why isn't it rational? Again we seem to have little choice. The explanation must cite whatever it is that we think prevents you from having knowledge. So, if we think that you do not know that you are sitting upon a chair because your belief is the product of perceptual equipment that is unreliable in your current environment, this same fact must be what precludes your belief from being rational. The more closely rational belief is tied to knowledge, the more difficult it is to avoid this conclusion.

My counterproposal is that the prerequisites of rational belief are not so closely tied to the conditions of knowledge. More exactly, the proposal is that this is so for the sense of rational belief that presupposes the egocentric perspective. This is not the only sense of rational belief. On the contrary, we evaluate beliefs from a variety of perspectives, depending on the context and our purposes, and we tend to give expression to these evaluations using the language of rationality.¹ The more objective the perspective that is presupposed, the more plausible will be the idea that a rational true belief absent Gettier problems is always an instance of knowledge. However, this is not so for egocentrically rational belief. The evil demon or the scientist who envats your brain deprive you of knowledge, but they need not deprive you of the opportunity of being egocentrically rational. This is the real lesson of the evil demon and the brain in the vat. By hypothesis these are situations that you could not distinguish from what you take to be your current situation. From your skin in, everything about these situations is as it is now. And yet, from your skin out, things are drastically different from what you take them to be in the current situation. Still, you would have egocentric reasons in such situations to believe exactly what you do now. The demon does not deprive you of these reasons. Rather, he alters your environment so that these reasons are no longer reliable indicators of truths. In so doing he deprives you of knowledge.

Knowledge, then, requires an element of luck, of good fortune. We cannot altogether eliminate the possibility of massive error by being egocentrically rational. We need the world to cooperate. This is what skeptical hypotheses teach us.² Knowledge is not within our control to the degree that egocentric rationality is. If contrary to what we think, the world or something in it conspires against us, then so much the worse for us as

knowing creatures. Nothing that we can do with respect to getting our own house in order will succeed in bringing us knowledge. This is not a comforting thought. We like to think of knowledge as part of our birthright. The thought that it might not be is so discomfiting that it makes an appeal to idealism in one of its many garbs attractive to some. This is an appeal to be resisted. It has all the advantages of metaphysics over a straightforward assessment of our situation. The better alternative is to give up success as a condition of egocentric rationality – to admit that this kind of rationality in and of itself is not enough to guarantee either truth or reliability.

Many of us will find it difficult to admit this, especially when doing philosophy. Among philosophers it is often taken for granted that the worst charge that we can make against others is that they are irrational. This attitude finds its way into our ethics as well as our epistemology. We resist the idea that egoists can be as rational as the rest of us. We think that we must prove that they are irrational, as if we would be at a loss as to how to criticize them if we could not do so. The remedy is to remind ourselves that not every failure need be one of rationality. There can be other explanations for moral failures. They might be the result of inadequate moral training, for example – a training that did not sufficiently develop our moral sensitivities. As a result, we might not be able to discriminate finely enough among the relevant features of morally difficult situations. Or more seriously, it may have left us with a fundamentally flawed character, one that has us caring for the wrong things.

Analogously, we may be tempted to think that someone who has massively mistaken beliefs must be irrational, as if this were the only possible explanation of their being so thoroughly misguided. But again, we need to remind ourselves that not every failure is a failure of rationality. There are other explanations for intellectual error, even widespread error. Like moral failure, it might be largely a matter of bad training. We might have been brought up in a culture whose intellectual traditions encourage error, a tradition that emphasizes magic, for example. Or more ominously, we might have inappropriate cognitive equipment. We might not be cognitively suited to detect truths in the environment in which we find ourselves. But whatever the explanation, the point is the same: Rationality in the theoretical sphere need be no more intimately tied to knowledge than

it is to goodness in the practical sphere. Just as you can be rational and yet lacking in virtue, so too you can be rational and yet lacking in knowledge. Appreciating this can help cure the preoccupation with skepticism that has dominated modern epistemology. It can allow egocentric epistemology to be done non-defensively.³

A non-defensive epistemology is one that refuses to apologize for a lack of guarantees. There is no guarantee that by being rational you will avoid error. There is no guarantee that you will avoid massive error. It need not even be probable that you will avoid massive error. Much of the implausibility of the Cartesian project arises from its failure to recognize that this is part of our intellectual condition. It instead insists that by being rational we can be assured of success. This insistence has disastrous consequences for egocentric epistemology. For contrary to what Descartes thought, there is nothing that we can do with respect to marshalling our cognitive resources that will result in such guarantees. Marshall them as we please. We will still need the world to cooperate. Consider the trust that we place in our perceptual equipment. If unbeknownst to us there is a deceiving demon in this world, then many of our perceptual beliefs will be false. And if most other close worlds are also demon worlds, then trusting our perceptual equipment does not even make it probable that we will avoid massive error.⁴ A non-defensive epistemology refuses to be intimidated by this possibility. It refuses to be intimidated into making success or likely success a prerequisite of rationality. It allows that it might be rational for us to trust our perceptual equipment even if doing so, unbeknownst to us, is likely to result in massive error.

It is a mistake for an egocentric epistemology to insist upon any kind of guarantee whatsoever between rationality and truth or likely-truth. This is the deepest flaw in the Cartesian approach to epistemology. It is not just that Descartes tried to guarantee too much, although this too is so. He unrealistically insisted that by being egocentrically rational we can be altogether assured of avoiding error. He was thus forced to regard any skeptical conjecture, no matter how far-fetched, as a *prima facie* defeater, one which itself had to be conclusively defeated before a claim could be rationally believed. But of course, if this were so, not much of anything would be rational for us to believe.

It might seem that the solution is simply to weaken the guarantee, but this would still leave

us with a defensive epistemology, and one that would face exactly the same problem that plagues Cartesian epistemology. This problem arises regardless of the strength of the guarantee, and it arises in exactly the same form as it did for Descartes. It arises if we say that by being rational we can be assured of having mostly true beliefs. It arises if we say more cautiously that by being rational we can at least be assured of avoiding the likelihood of massive error. It even arises if we say that by being rational we can be assured only that the likelihood of our avoiding error is greater than if we were not rational.⁵ For regardless of the nature of the guarantee, there will be no non-question begging assurances that the way in which we are marshalling our cognitive resources generates beliefs that meet the guarantee. There will be no non-question begging assurances, in other words, that the way in which we are marshalling our resources is suitable for our environment.

After all, the search for such assurances will itself require us to marshal our cognitive resources. It will itself involve the use of methods about which we can sensibly have doubts, doubts that cannot be addressed without begging the question. Any attempt to address them will employ methods either that are themselves already at issue or that can be made so. There is a close analogy with the practical realm. There too self-directed inquiry can raise doubts that cannot be addressed without begging the question. I commit myself to various projects, ones that initially seem worthwhile, but if I examine my commitments and the values implicit in them, doubts can occur to me. I can ask whether I want to be the kind of person who makes these sorts of commitments. Can I endorse my being that kind of person? And even if I answer "yes," this does not definitively settle the doubts. I can go on and ask about the values implicit in this latest endorsement. Either they are values that were implicit in the commitments about which I originally had doubts or they are new values about which I can also raise doubts. It is hopeless to search for a non-question begging way to endorse all of our values, including the values implicit in the endorsement itself. Any search of this sort would be based on the assumption that there is a neutral position from which such endorsements can be made, but there isn't. Nor is there in epistemology. There is no neutral position from which to defend our intellectual commitments.

But if not, we must admit that egocentric epistemology cannot provide non-question begging assurances that we will avoid massive error by being rational. The search for such assurances is doomed from the start. It is one thing to insist that skeptical hypotheses are genuinely possible. It is another to insist that the rationality of our beliefs depends upon our having a non-question begging way to discharge them. We have no such way, and our rationality does not depend upon our having one.

Admitting this need not lead to quietism. One of our intellectual projects, arguably our most fundamental one, is to understand our own position in the world, including our position as inquirers. Within the context of such a project, it is natural to raise general doubts about our intellectual commitments. It is natural to entertain even radically skeptical doubts about them. Of course, making ourselves into an object of systematic inquiry is not an everyday occurrence. It requires some detachment from our ordinary concerns. You cannot with sanity raise general questions about your intellectual commitments when, say, discussing with your mechanic the problems you are having with your car.⁶ Nor can you raise them when you are doing physics or biology or geometry. But in the context of an inquiry into our place in the world, they arise without force. We make ourselves into the objects of our study, and we recognize that these objects that we are studying are creatures who have a rich interaction with their environment. They have various beliefs about it and various desires for it, all of which become intertwined in their projects. The intellectual projects that find expression in their sciences, for example, are intertwined with projects that are aimed at controlling their environment. These projects, we further recognize, can be conducted more or less successfully. In wondering about the relative success of their intellectual projects, we are raising general questions about their beliefs, questions that make it natural to entertain skeptical hypotheses. We are wondering whether their cognitive equipment and their ways of employing this equipment are sufficiently well-suited for their environment as to be prone to produce true beliefs about it. Even in wondering about the success of their non-intellectual projects, these same questions arise indirectly. For even if we grant that these creatures are mostly successful in controlling their environment, it is natural to want some explanation for this success. Is it by having largely accurate beliefs about their

environment that they are able to exercise this control or is there some other explanation? But in wondering whether there might not be another explanation, we are once again taking skeptical possibilities seriously. It is perfectly natural for us to do so in this context.

So, it is not a mistaken philosophical tradition that leads us to skeptical thoughts.⁷ It is our natural curiosity. We are curious about these creatures' place in the world, including their place as inquirers. We know that they take themselves to be accurately representing their world. It is natural for us to wonder whether they are right in thinking this or whether their representations might be distorted in some systematic way. The hypothesis of the evil demon and the brain in the vat are merely dramatic devices to express these kinds of thoughts in their most radical form.

There is, of course, something else that is unusual about these kinds of thoughts. They are about our beliefs, our presuppositions, our methods of inquiry. If we are to make these things the objects of concern, we must be able to distance ourselves from them in some way. This might make it seem as if the entertainment of skeptical hypotheses is inevitably an exercise in schizophrenia.

But if this be schizophrenia, it is of a common enough sort. Indeed, it is easy to come by even in the limiting case of belief, that which is indubitable for you. Such propositions are irresistible for you once you bring them clearly to mind. Clarity about them is enough to command your assent.⁸ So, you cannot directly doubt the truth of such a proposition. I may be able to do that, but you cannot. Otherwise, it would not be genuinely indubitable for you. Even so, you can do the next best thing. You can raise questions about its truth indirectly. You can do so by considering in a general way whether that which is indubitable for you is really true. You can wonder whether you might not be the kind of creature who finds certain falsehoods impossible to doubt. Your wondering this does not prove that nothing is really indubitable for you. It does not prove that you really are capable of doubting that you exist and that $2 + 1 = 3$. These propositions can still be irresistible for you whenever you directly consider them. However, you can refuse to do this. You can refuse to bring them fully to mind, and by so refusing you gain the ability to suspend belief in them hypothetically. You need not cease believing them. You merely cease focusing your direct attention upon them. In doing this you can distance yourself even from that

which is indubitable for you, and thus you can make even these propositions an object of skeptical concern. There is nothing mysterious about your doing so.⁹ Similarly, there is nothing mysterious about your entertaining serious and general skeptical worries about the other propositions that you believe. You can doubt in a general way whether much of what of you believe is true, and you can do so without actually giving up those beliefs. It is enough for you to suspend them hypothetically.

To think that there is something inevitably puzzling about entertaining general skeptical doubts is to make human thought into something far less flexible than it is. Atheists can debate even theological questions with theists and they can do so without altering their beliefs. They hypothetically suspend for the duration of the discussion a good portion of what they believe. Similarly, the morally upright can appreciate, admire, and even enjoy the ingenuity and resourcefulness of literary villains even when that ingenuity and resourcefulness is put to repugnant purposes. They can do so by hypothetically suspending their moral scruples. There may be limits as to how much of our beliefs and values we can put into suspension, but the limits are at best distant ones. They are not so constraining as to prevent sensible discussion between atheists and theists, and they are not such as to preclude appreciation of the great literary villains. Nor are they so stringent as to rule out worries about the general reliability of our beliefs. We need not abandon our beliefs in order to entertain such worries. It is enough for us to suspend them hypothetically.¹⁰

The way to respond to skeptical doubts is not to legislate against them metaphysically, and it is not to dismiss them as meaningless, self-defeating, or even odd. It is rather to live with them. It is to recognize that what makes epistemology possible also makes skeptical worries inevitable – namely, our ability to make our methods of inquiry themselves into an object of inquiry. Within the context of such an inquiry, the worry that we might possibly have widely mistaken beliefs is as natural as it is ineradicable. If this illustrates our whimsical condition, then so be it,¹¹ but it is, after all, not so surprising. We want to be able to defend or at least explain the reliability of our methods of inquiry, but the only way to do so is within our own system of inquiry. We seek to use our methods to show that these same methods are to be trusted. This leaves us vulnerable to the charge that we are begging the question against the skept-

tic. If the only way to defend or to explain our general way of proceeding is by using that way of proceeding, then we will not have altogether ruled out the possibility that its products might be widely mistaken. This is no more than a generalization of the problem of the Cartesian circle, and it is a circle from which we can no more escape than could Descartes.

But if we too are caught in this circle, what's the point of inquiring into the reliability of our methods of inquiry? Why not relax and just assume that our fundamental methods are reliable? Why not encourage or at least tolerate intellectual complacency about these matters? Because striving to use our fundamental methods of inquiry to defend or to explain their own reliability is far from pointless. Besides, even if it were, it would not matter. We cannot help ourselves. Our curiosity compels us to seek such explanations. But in fact, it is not pointless to seek them, since they need not always be forthcoming. Not all methods of inquiry are even capable of begging the question in their own defense, even though this is the least we should expect from them. The least we should expect is that they be self-referentially defensible. What this means, for beginners, is that they be logically coherent. It must be possible to employ them in their own defense,¹² but possibility is not enough. In addition, the circumstances have to be favorable for such a defense, and the methods themselves might indicate that this is not so. It is sometimes suggested that the collection of procedures that we call "the scientific method" is self-referentially indefensible in just this way. The history of science, it is argued, is largely a history of error. We look back at the theories of even the best scientists of previous times and find much in those theories that is false and even silly. Moreover, there is no reason to think that future scientists won't think the same of our best current theories. In this way, the history of science might seem to give us good inductive grounds – grounds that are themselves acceptable given the methods of science – for thinking that the scientific method is unreliable.

This is a perfectly respectable argumentative strategy. If the use of the scientific method in our environment has been proven, in accordance with canons acceptable to that method, to generate mistaken theories with regularity, then so much the worse for it as a procedure to generate true theories. The least we should expect from a proposed method of inquiry is that it be able to defend itself

in its own terms. Much of recent philosophy of science can be read as trying to do just that. It can be read, that is, as trying to give a construal of the scientific method and a reading of the history of science that together constitute a response to this pessimistic induction. For example, there are those who claim that any fair look at the history of science reveals not so much a history of repudiation of past theories but rather a history in which past theories are largely incorporated into their successor theories. In addition, they claim that the immediate aim of scientific theorizing is not so much to generate theories that are strictly and thoroughly true but rather ones that are at least approximately true. The aim is verisimilitude. They then point out that the history of science, so understood, provides no basis for an induction whose conclusion is that present theories are not even approximately true. On the contrary, that history is marked by ever increasing predictive success, and the best explanation of this is that those sciences are getting closer and closer to the truth. So, far from supporting a pessimistic induction, the history of science gives us good reason to think that the terms of our sciences, especially our more mature ones, typically do refer.¹³

It is tempting to dismiss arguments of this sort on the grounds that they beg the question. After all, the scientific method is a method that makes essential use of arguments to the best explanation. So, questions about its reliability are in large measure questions about the truth preservingness of such arguments. And yet, the response employs an argument to the best explanation in order to defend the scientific method. It is thus presupposing exactly that which it is trying to establish.

Even so, what I have been arguing is that some questions deserve to be begged. Questions about the reliability of our fundamental methods of inquiry are just such questions. It need not be a fault of the scientific method that it cannot be defended without begging the question. The fault would lie in there being no argument by which the method can be defended. If there is no way that the method can be defended, not even a question begging way, then it would fail even the minimum test for a method of inquiry. This is only the minimum, however. There are patently silly methods that can be used to defend themselves.¹⁴ So, if the only thing that can be said in favor of the scientific method is that it can be used to defend itself, this is not much. It is certainly not enough to provide assurances of its reliability, as proponents

of arguments to the best explanation sometimes hint. On the other hand, it is not altogether insignificant either, as their opponents sometimes hint.

Likewise, it is misguided to complain about the Cartesian circle, not because Descartes did not argue in a circle – he did – but rather because this is not the flaw in his strategy. The problem is not that he begs the question by appealing to what he takes to be clear and distinct considerations in order to show that clarity and distinctness assures us of truth. If a proposed method of rational inquiry is fundamental, it cannot help but be used in its own defense if it is to be defended at all. The problem, rather, is that Descartes thought that his strategy, if successful, could altogether extinguish serious skeptical worries. He was wrong about this. Suppose that Descartes had in fact provided a clear, distinct and hence irresistible proof of God's existence and had succeeded also in providing an irresistible proof that God would not allow that which is irresistible for us to be mistaken. This still would not have been enough to answer all of the skeptic's questions, although admittedly it perhaps would come as close as possible to doing so. In large part it is this that makes the Cartesian strategy such an appealing one. If the arguments work, would-be skeptics are forced to go to extreme lengths to keep their skeptical concerns alive, but they can do so. They will not be able to do so as long as they have Descartes' irresistible proofs clearly in mind, for as long as these proofs are clearly in mind even would-be skeptics cannot help but believe that irresistible propositions are true. But of course, they need not always have the proofs in mind. Thus, as with other propositions, they can suspend belief hypothetically in the proposition that irresistible propositions are true. They can distance themselves from the spell of the proofs' irresistibility and by so doing they can sensibly raise the question of whether irresistibility really is sufficient for truth. Descartes can urge them to recall his proofs, since by hypothesis this will dispel all of their doubts. However, and this is the most important point, while not under the influence of the irresistible proofs, the would-be skeptics can grant that recalling the proofs would have this effect upon them and yet still insist that this does not settle the issue of whether irresistibility really is sufficient for truth. And they would be right.¹⁵

Thus, there is nothing wrong with trying to appeal to clear and distinct and hence irresistible ideas in an attempt to argue that such ideas are

true. One of the things we should expect of those proposing strategies of inquiry is that they be able to use these strategies to defend their own proposals. On the other hand, it is a mistake to think any such defense will be capable of altogether eliminating skeptical worries. Skeptical worries are ineradicable, and we might as well get ourselves used to this idea.

Doing so will involve admitting that it is alright to do epistemology, and egocentric epistemology in particular, non-defensively. The prerequisite of egocentric rationality is not truth or even reliability but rather the absence of any internal motivation for either retraction or supplementation of our beliefs. Egocentric rationality requires that we have beliefs that are to our own deep intellectual satisfaction – ones that do not merely satisfy us in a superficial way but that would do so even with the deepest reflection. So, to be egocentrically rational is to be invulnerable to a certain kind of self-condemnation. It is to have beliefs that in our role as truth-seekers we wouldn't criticize ourselves for having even if we were to be deeply reflective. There are various ways of trying to say what exactly this amounts to,¹⁶ but for the issue at hand these details are not important. What is important is that even if we are deeply satisfied with our beliefs, we cannot be assured of avoiding massive error. There are no such assurances. There are not even assurances of it being even likely that we will avoid massive error. The lack of assurances is built into us, and it is built into the nature of our inquiries. We must do epistemology with this in mind.

Even so, it is equally important to remember that being deeply satisfied with one's methods and beliefs is not everything. You might be deeply satisfied with them because you are dogmatic, for example. You might have views about your methods of inquiry that effectively protect them against self-directed challenges. You have ready explanations for any would-be oddity in your method or in the beliefs that they generate. Take astrology as a case in point. Most contemporary astrologers may be impostors, but suppose you are not. You are deeply convinced of its truth. No amount of disinterested reflection would prompt you to be critical of your methods or of the beliefs that they produce. Are your beliefs irrational? Not necessarily. Are they dogmatic and misguided? Of course.

Most of us are not afflicted with this kind of extreme dogmatism. If we are dogmatic, we are unlikely to be so all the way down. The deepest epistemic standards of even the most dedicated

astrologers are not likely to be radically different from those of the rest of us. But if so, they are likely to be vulnerable to self-criticism. They themselves would be suspicious of their methods and their beliefs were they to be sufficiently impartial and sufficiently reflective.

There are those, no doubt, who will find this naive. Perhaps it is. But if so, the alternative is not to make all astrologers irrational by fiat. It is rather to admit that some might be rational albeit fundamentally misguided. The impulse to inject every intellectual desirable characteristic into the theory of rationality is one to be resisted. It is not inconceivable that someone can be dogmatic without being irrational. My approach is to explain as much dogmatism as possible internally. It is to rely upon our own characters as inquirers. Most dogmatists, I claim, are violating their own deepest standards. If there are some dogmatists left over, some who are not violating even their own deepest standards, they are to be dismissed as dogmatic and that is the end of the matter. It is a mistake to try to construct an objective theory of dogmatism and then to make the avoidance of that kind of dogmatism a prerequisite of rationality. Not every shortcoming is one of rationality.

Again, there is a useful analogy between the practical and the intellectual. There is no more unity among intellectually desirable characteristics than there is among non-intellectual ones. Our actions are egocentrically rational insofar as we lack internal motivations to be dissatisfied with them. Much immoral behavior can be criticized as being irrational in just this way. We do what we ourselves cannot sincerely endorse, given our own deepest values. But of course, this makes the irrationality of immorality contingent upon our characters. It makes it contingent upon our deepest values. If there are fanatics who lack even a deeply internal motivation to detach from their vicious behavior, then we must be content with regarding them as fanatics. Their problem, and ours, is that they have vicious characters. They need not be irrational.

This is not to say that there is not a looser kind of unity between egocentric rationality and morally desirable characteristics on the one hand and between egocentric rationality and intellectually desirable ones on the other. It may be that when we have no internal motivations, not even deep ones, to be dissatisfied with what we are doing, then in general we will be acting in a morally virtuous way, and it may be that when we are not

acting virtuously, there generally will be some internal motivation for detachment that we have ignored or not noticed. Our normal psychological make-up may ensure that in general this is so. As a result, it may be that egocentric rationality and morality go hand-in-hand except in situations that are bizarre or in people who are deranged.

Similarly for egocentric rationality and intellectually desirable characteristics. It may be that when we have no internal motivations to retract what we believe, then in general we are neither dogmatic nor thoroughly misguided. It likewise may be that when we are either dogmatic or misguided, there is in general some internal motivation for retraction that we have ignored. Thus, it may be that egocentric rationality and knowledge, like egocentric rationality and morality, go hand-in-hand except in situations that are bizarre or in people who are deranged. If so, it will also be the case that except in such situations or with such people, we can expect disagreements of opinion to be largely the result of differences of information. Or put the other way around, if our intellectual peers persist in disagreeing with our opinions despite the fact that they have access to the same information, this calls for some explanation. It won't do simply to say that they are wrong and we are right. On the contrary, unless there is some plausible way to explain away their opinion, the disagreement ordinarily will give us a good reason to be suspicious of our own opinions.

All these things may well be so. The mistake is to try to make these general truths, if that be what they are, into categorical ones. It is a mistake to make it a matter of necessity that being egocentrically rational is likely to bring us knowledge, and it is equally a mistake to assume that those who have fundamentally misguided beliefs – even those who are misguided to the point of being deranged – must of necessity also be irrational. Correspondingly, it is a mistake to make it a matter of necessity

that rational people will agree with one another if they have the same information. One of the presuppositions of the Cartesian project was that rationality is what stands between us and “a chaotic disagreement in which anything goes.”¹⁷ However, this need not be our presupposition. We can say that what stands in the way of chaotic disagreement is not simply the nature of rationality but also the contingent fact that we are born with similar cognitive equipment and into similar environments, a contingent fact that makes it likely that the deep epistemic standards of one person will not be radically different from those of another.

This then is a sketch of a way to think about egocentric rationality. According to this conception, egocentric rationality brings with it no guarantees of truth or likely-truth, and as a result it brings with it no guarantees that rational people with access to the same information will agree with one another. Why, then, should we be interested in egocentric rationality? Because we are interested in having beliefs that are accurate and comprehensive and because by being egocentrically rational we will be pursuing this end in a way that by our own lights seems effective. To be sure, this involves a leap of intellectual faith. It involves our having confidence in those intellectual methods that are deeply satisfying to us despite the fact that we cannot vindicate this confidence in a non-question begging way. This may be regrettable but it is also undeniable. The reality of our intellectual lives is that we are working without nets. No procedure, no amount of reflection, no amount of evidence gathering can guarantee that we won't fall into error, perhaps even massive error. We are thus forced to choose between proceeding in a way that we on reflection would take to be effective and proceeding in a way that we would not take to be effective. If we are rational, we opt for the former.

Notes

- 1 See Richard Foley, *The Theory of Epistemic Rationality* (Cambridge: Harvard University Press, 1987), especially sec. 2.8.
- 2 In his book *Philosophical Explanations* (Cambridge: Harvard University Press, 1981), Robert Nozick insists that for a belief to constitute knowledge, it must be non-accidentally true. At first glance this might seem to be at odds with what I am

claiming, but in fact it isn't. Knowledge may very well require that truth and belief be non-accidentally related, so that given the belief that *P* it is not matter of luck that *P* is true, and vice-versa. Even so, we need an element of luck – we need the world to cooperate – in order for there to be this non-accidental relationship between truth and belief.

- 3 I have borrowed the phrase 'defensive epistemology' from Bas van Fraassen.
- 4 I assume here that if p is probable given q , then it cannot be the case that p is false in most close situations in which q is true.
- 5 Socrates argued that one never acts for the worse by having a virtue. The claim here is analogous; one never believes for the worse by being rational, where believing for the worse is a matter of believing in a way that is more likely to lead to error. Rationality, it is claimed, guarantees at least this much.
- 6 Descartes himself emphasized this, insisting that his method of doubt is not appropriate for use in ordinary life. See his *Discourse on Method*, in Descartes, *Philosophical Writings*, vol. I, eds. Haldane and Ross (Cambridge: Cambridge University Press, 1911), especially pp. 100–1.
- 7 See Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979).
- 8 In the *Fifth Meditation*, Descartes says "... the nature of my mind is such that I would be unable not to assent to these things (which I clearly and distinctly perceive) so long as I clearly perceive them..." See Haldane and Ross, vol. I, p. 180.
- 9 Compare with Bernard Williams, *Descartes: The Project of Pure Inquiry* (Harmondsworth: Penguin Books, 1978), and "Descartes' Use of Skepticism," in M. Burnyeat (ed.), *The Skeptical Tradition* (Berkeley: University of California Press, 1983), pp. 337–52. See also Anthony Kenny's discussion of first-order doubt and second-order doubt in "The Cartesian Circle and the Eternal Truths," *The Journal of Philosophy* 67(1970), pp. 685–700; and James Van Cleve, "Foundationalism, Epistemic Principles, and the Cartesian Circle," this vol., ch. 20.
- 10 Contrast this with Hume, whose position does hint of schizophrenia. Hume reported that while engaged in philosophical reflection he found himself forced to give up his ordinary beliefs about material objects, the future and the self, but that as soon as he left his study these reflections appeared strained to him and his ordinary beliefs returned. So, according to Hume, his beliefs changed dramatically depending upon which of his personalities was engaged, his philosophical one or his everyday one. The potential for schizophrenia here is made even more dramatic if we assume not only that the philosophical Hume knew that were he to quit reflecting he would begin believing but also that the everyday Hume knew (at least tacitly) that were he to begin reflecting he would cease believing.
- 11 "When [a Pyrrhonian] awakes from his dream, he will be the first to join in the laugh against himself, and to confess, that all his objections are mere amusement, and can have no other tendency than to show the whimsical condition of mankind, who must act and reason and believe; though they are not able, by their most diligent enquiry, to satisfy themselves concerning the foundations of these operations, or to remove the objections, which may be raised against them." David Hume, *An Enquiry Concerning Human Understanding*, ed. L. A. Selby-Bigge with text revised by P. H. Nidditch (Oxford: Oxford University Press, 1975), sec. XII, p. 128.
- 12 Alvin Plantinga claims that certain versions of classical foundationalism cannot possibly be used to defend themselves. See Plantinga, "Is Belief in God Rational?" in C. F. Delaney (ed.), *Rationality and Religious Belief* (Notre Dame: University of Notre Dame Press, 1979), pp. 7–27.
- 13 See, e.g. Hilary Putnam, *Reason, Truth and History* (Cambridge: Cambridge University Press, 1986); William Newton-Smith, *The Rationality of Science* (London: Routledge & Kegan Paul, 1981); Ernan McMullin, "The Fertility of Theory and the Unity of Appraisal in Science," in R. S. Cohen *et al.* (eds), *Boston Studies* 39 (1976), pp. 395–432.
- 14 Ernest Sosa gives the following example: "If a rule or principle contains the proposition that the earth is flat, then it is acceptable, as is the proposition that the earth is flat." See Sosa, "Methodology and Apt Belief," *Synthese* 74 (1988), p. 418.
- 15 Contrast with Bernard Williams: "So the believer can always recall the skeptic, unless the skeptic is willfully obstinate, to considering the existence and benevolence of God, and if the skeptic concentrates on those proofs, he will believe not only those propositions themselves but also something that follows from them – namely, that clear and distinct perceptions are reliable, and hence skepticism unjustified." Williams, "Descartes' Use of Skepticism," p. 349. My position, in turn, is that in this context, where the issue is precisely whether irresistible proofs might be mistaken, it is not mere obstinacy to refuse to recall the proofs. The skeptic's refusal to recall them is not unlike the alcoholic's refusal to enter the bar. In each case the refusal is motivated by a fear that one's weaknesses will be exploited.
- 16 For my suggestions, see *The Theory of Epistemic Rationality*.
- 17 The phrase is from Bernard Williams. See Williams, "Descartes' Use of Skepticism," p. 344.

Epistemic Norms

John Pollock

Epistemic Norms

What are we asking when we ask whether a belief is justified? What we want to know is whether it is all right to believe it. Justification is a matter of "epistemic permissibility." It is this normative character of epistemic justification that I want to emphasize. That epistemic justification is a normative notion is not a novel observation. The language of epistemic justification is explicitly normative, and a recurrent theme has been that justification is connected with the "ethics of belief." This has played a role in the thought of a number of epistemologists: Chisholm (1977 and chapter one of 1957) has repeatedly stressed the normative character of epistemic terms, several recent philosophers have proposed analyzing epistemic justification in terms of the maximization of epistemic values,¹ and a few philosophers have appealed to the normative character of justification in other ways.² Thus I will think of epistemic justification as being concerned with questions of the form, "When is it permissible (from an epistemological point of view) to believe P?" This is the concept of epistemic justification that I am concerned to explore.

Norms are general descriptions of the circumstances under which various kinds of normative judgments are correct. Epistemic norms are norms describing when it is epistemically permissible to hold various beliefs. A belief is justified if and only if it is licensed by correct epistemic

norms. We assess the justifiedness of a belief in terms of the cognizer's reasons for holding it, and our most fundamental epistemic judgments pertain to reasoning (construing reasoning in the broad manner required by direct realism). Thus we can regard epistemic norms as the norms governing "right reasoning." Epistemic norms are supposed to guide us in reasoning and thereby in forming beliefs. The concept of epistemic justification can be explained by explaining the nature and origin of the epistemic norms that govern our reasoning. I have called this "the reason-guiding concept of epistemic justification." There may be other concepts that can reasonably be labeled "epistemic justification," but it is the reason-guiding concept that is the focus of the present chapter and is involved in traditional epistemological problems.

Much of recent epistemology has been concerned with describing the contents of our epistemic norms, but the nature and source of epistemic norms has not received much attention. Epistemologists have commonly supposed that epistemic norms are much like moral norms and that they are used in evaluating reasoning in the same way moral norms are used in evaluating actions. One of the main contentions of this chapter will be that this parallel is not at all exact and that epistemologists have been misled in important ways by supposing the analogy to be better than it is. A proper understanding of epistemic norms will provide us with a radically new perspective on epistemology, and from the point of view of this perspective new light can be thrown on a number of central epistemological problems.

Originally published in J. Pollock, *Contemporary Theories of Knowledge* (Lanham: Rowman and Littlefield, 1986), ch. 5.

An account of epistemic norms must answer two different questions. First, it must describe the correct epistemic norms. Second, it must tell us what makes them correct. The first question concerns the content of epistemic norms, and the second question concerns their justification. By distinguishing between these questions we can see the internalism/externalism distinction in a new light. A belief is justified if and only if it is held in conformance with correct epistemic norms. Externalism is the view that the justifiedness of a belief is a function in part of external considerations. Thus if externalism is right, external considerations must play a role in determining whether a belief is held in conformance with correct epistemic norms. This could arise in either of two ways. On the one hand, external considerations could enter into the formulation of correct epistemic norms. On the other hand, it might be granted that epistemic norms can only appeal to internal considerations, but it might be insisted that external considerations are relevant to determining which set of internalist norms is correct. Thus we are led to a distinction between two kinds of externalism. *Belief externalism* insists that correct epistemic norms must be formulated in terms of external considerations. A typical example of such a proposed norm might be "It is permissible to hold a belief if it is generated by a reliable cognitive process." In contrast to this, *norm externalism* acknowledges that the content of our epistemic norms must be internalist, but employs external considerations in the selection of the norms themselves. The distinction between belief and norm externalism is analogous to the distinction between act and rule utilitarianism. Externalism (simpliciter) is the disjunction of belief externalism and norm externalism. A number of philosophers who are usually considered externalists appear to vacillate between belief externalism and norm externalism. The difference between these two varieties of externalism will prove important. In the end, both must be rejected, but they are subject to different difficulties.³

According to internalism, the justifiedness of a belief is a function exclusively of internal considerations, so internalism implies the denial of both belief and norm externalism. That is, the internalist maintains that epistemic norms must be formulated in terms of relations between beliefs or between beliefs and nondoxastic internal states (e.g., perceptual states), and he denies that these norms are subject to evaluation in terms of external

considerations. Typically, the internalist has held that whatever our *actual* epistemic norms are, they are necessarily correct and not subject to criticism on any grounds (externalist or otherwise), but I have not built that into the definition of internalism.

There is one respect in which the internalism/externalism distinction remains to be made clear. The distinction is formulated in terms of an undefined notion of an internal state. It is fairly clear what kinds of states people have had in mind when they have talked about internalism and externalism, but it is hard to give a general characterization of them. I will return to this matter in section four.

How Do Epistemic Norms Regulate?

In order to get a grasp of the nature of epistemic norms, let us begin by asking their purpose. It is important to distinguish between two uses of norms (epistemic or otherwise). On the one hand, there are third-person uses of norms wherein we use the norms to evaluate the behavior of others. Various norms may be appropriate for third-person evaluations, depending upon the purpose we have in making the evaluations. For example, we may want to determine whether a person is a good scientist because we are trying to decide whether to hire him. To be contrasted with third-person uses of norms are first-person uses. First-person uses of norms are, roughly speaking, action-guiding.⁴ For example, I might appeal to *Fowler's Modern English Usage* to decide whether to use "that" or "which" in a sentence. Epistemological questions are inherently first-person. The traditional epistemologist asks, "How is it possible for me to be justified in my beliefs about the external world, about other minds, about the past, and so on?" These are questions about what to believe. Epistemic norms are the norms in terms of which these questions are to be answered, so these norms are used in a first-person reason-guiding capacity.

If reasoning is governed by epistemic norms, just how is it governed? There is a model of this regulative process that is often implicit in epistemological thinking, but when we make the model explicit it is *obviously* wrong. This model assimilates the functioning of epistemic norms to the functioning of explicitly articulated norms. For example, naval officers are supposed to "do it by the book," which means that whenever they are in doubt about what to do in a particular situation

they are supposed to consult explicit regulations governing all aspects of their behavior and act accordingly. Explicitly articulated norms are also found in driving manuals, etiquette books, and so on. Without giving the matter much thought, there is a tendency to suppose that all norms work this way, and in particular to suppose that this is the way epistemic norms work. I will call this “the intellectualist model.”⁵ It takes little reflection to realize that epistemic norms cannot function in accordance with the intellectualist model. If we had to make an explicit appeal to epistemic norms in order to acquire justified beliefs we would find ourselves in an infinite regress, because to apply explicitly formulated norms we must first acquire justified beliefs about how they apply to this particular case. For example, if we are to reason by making explicit appeal to a norm telling us that it is permissible to move from the belief that something looks red to us to the belief that it is red, we would first have to become justified in believing that that norm is included among our epistemic norms and we would have to become justified in believing that we believe that the object looks red to us. In order to become justified in holding those beliefs, we would have to apply other epistemic norms, and so on *ad infinitum*. Thus it is clear that epistemic norms cannot guide our reasoning in this way.⁶

If the intellectualist model is wrong, then how do epistemic norms govern reasoning? At this point we might raise the possibility that they do not. Perhaps epistemic norms are only of use in third-person evaluations. But it cannot really be true that epistemic norms play *no role at all* in first-person deliberations. We can certainly subject our reasoning to self-criticism. Every philosopher has detected invalid arguments in his own reasoning. This might suggest that epistemic norms are only relevant in a negative way. Our reasoning is innocent until proven guilty. We can use reasoning to criticize reasoning, and hence we can use reasoning in applying epistemic norms to other reasoning, but we cannot be required to reason about norms *before* we can do any reasoning. This would avoid the infinite regress.

But as theoretically attractive as the “innocent until proven guilty” picture might be, it cannot be right. It entails the view, that I have already rejected elsewhere, according to which all beliefs are *prima facie* justified. This view cannot handle the fact that epistemic norms guide the acquisition of beliefs and not just their after-the-fact evalua-

tion. Even in the perceptual acquisition of beliefs about physical objects, the resulting beliefs are sometimes unjustified. More generally, there are a number of natural processes that lead to belief formation. Among these are such “approved” processes as vision, inductive reasoning, deductive reasoning, and memory, and also some “unapproved” but equally natural processes such as wishful thinking. The latter is just as natural as the former. An example that I have used previously runs as follows. My daughter had gone to a football game, the evening had turned cold, and I was worried about whether she took a coat. I found myself thinking, “Oh, I am sure she is wearing a coat.” But then on reflection I decided that I had no reason to believe that – my initial belief was just a matter of wishful thinking. The point here is that wishful thinking is a natural belief-forming process, but we do not accord it the same status as some other belief-forming processes like vision. Although we have a natural tendency to form beliefs by wishful thinking, we also seem to “naturally” know better. This is not just a matter of after-the-fact criticism. We know better than to indulge in wishful thinking at the very time we do it. It seems that *while* we are reasoning we are being guided by epistemic norms that preclude wishful thinking but permit belief formation based upon perception, induction, and so on. This is of more than casual significance, because it might be impossible to rule out wishful thinking by after-the-fact reasoning. This is because the after-the-fact reasoning might include wishful thinking again, and the new wishful thinking could legitimize the earlier wishful thinking. If epistemic norms play no regulative role in our reasoning while it is going on, there is no reason to think they will be able to play a successful corrective role in after-the-fact evaluations of reasoning. In order for the corrective reasoning to be successful it must itself be normatively correct. Epistemic norms must, and apparently do, play a role in guiding our epistemic behavior at the very time it is occurring. But how can they?

Epistemic norms cannot play a merely negative, corrective, role in guiding reasoning, nor can they function in a way that requires us to already make judgments before we can make judgments. What is left? I think that our perplexity reflects an inadequate understanding of the way action-guiding norms usually function. The case of making an explicit appeal to norms in order to decide what to do is the exception rather than the rule. You

may make reference to a driving manual when you are first learning to drive a car, but once you learn how to drive a car you do not look things up in the manual anymore. You do not usually give any explicit thought to what to do – you just do it. This does not mean, however, that your behavior is no longer guided by those norms you learned when you first learned to drive. Similarly, when you first learned to ride a bicycle you were told to turn the handlebars to the right when the bicycle leaned to the right. You learned to ride in accordance with that norm, and that norm still governs your bike-riding behavior but you no longer have to think about it. The point here is that *norms can govern your behavior without your having to think about them*. The intellectualist model of the way norms guide behavior is almost always wrong. This is an obvious point, but it has been insufficiently appreciated. It is of major importance in understanding epistemic norms. Reasoning is more like riding a bicycle than it is like being in the navy.

What makes it possible for your bike-riding behavior to be governed by norms without your thinking about the norms is that you *know how* to ride a bicycle. This is *procedural knowledge* rather than *declarative knowledge*. Having procedural knowledge of what to do under various circumstances does not involve being able to give a general description of what we should do under those circumstances. This is the familiar observation that knowing how to ride a bicycle does not automatically enable one to write a treatise on bicycle riding. This is true for two different reasons. First, knowing how to ride a bicycle requires us to know what to do in each situation *as it arises*, but it does not require us to be able to say what we should do before the fact. Second, even when a situation has actually arisen, our knowing what to do in that situation need not be propositional knowledge. In the case of knowing that we should turn the handlebars to the right when the bicycle leans to the right, it is plausible to suppose that most bicycle riders do have propositional knowledge of this; but consider knowing how to hit a tennis ball with a tennis racket. I know how to do it – as the situation unfolds, at each instant I know what to do – but even at that instant I cannot give a description of what I should do. Knowing what to do is the same thing as knowing to do it, and that need not involve propositional knowledge.

We are now in a position to give a rough explanation of how action-guiding norms can

govern behavior in a non-intellectualist manner. When we learn how to do something X, we “acquire” a plan of how to do it. That plan might (but need not) start out as explicit propositional knowledge of what to do under various circumstances, but then the plan becomes internalized. Using a computer metaphor, psychologists sometimes talk about procedural knowledge “being compiled.” When we subsequently undertake to do X, our behavior is automatically channeled into that plan. This is just a fact of psychology. We form habits or conditioned reflexes. Norms for doing X constitute a description of this plan for doing X. The sense in which the norms guide our behavior in doing X is that the norms describe the way in which, once we have learned how to do X, our behavior is automatically channeled in undertaking to do X. The norms are not, however, just descriptions of what we do. Rather, they are descriptions of what we *try* to do. Norms can be hard to follow and we follow them with varying degrees of success. Think for example, of an expert golfer who knows how to swing a golf club. Nevertheless, he does not always get his stroke right. It is noteworthy, and it will be important later, that when he does not get his stroke right he is often able to tell that by something akin to introspection. When he does it wrong it “feels wrong.” The ability to tell in this way whether one is doing something right is particularly important for those skills governing performances (like golf swings) that take place over more than just an instant of time, because it enables us to correct or fine tune our performance as we go along. The distinction between knowing how to do something and actually doing it is the same as the competence/performance distinction in linguistics. Our linguistic knowledge is procedural knowledge. We know how to use language, but we do not always use language correctly.

The internalization of norms results in our having “automatic” procedural knowledge that enables us to do something without having to think about how to do it. It is this process that I am calling “being guided by the norm without having to think about the norm.” This may be a slightly misleading way of talking, because it suggests that somewhere in our heads there is a mental representation of the norm and that mental representation is doing the guiding. Perhaps it would be less misleading to say that our behavior is being guided by our procedural knowledge and the way in which it is being guided is described by the

norm. What is important is that this is a particular way of being guided. It involves nonintellectual psychological mechanisms that both guide and correct (or fine tune) our behavior.

What we know in knowing how to ride a bicycle can be given a normative description. This procedural knowledge consists of knowing what to do under various circumstances, e.g., knowing to turn right when the bike leans to the right. This can equally be described as knowing what we *should* do under those circumstances. The point of using normative language to describe internalized norms is to contrast what the norms tell us to do with what we *do*. The simple fact of the matter is that even when we know how to do something (e.g., swing a golf club) we do not always succeed in following our norms. This use of “should” in describing procedural knowledge is interesting. Moral philosophers have talked about different senses of “should,” distinguishing particularly between moral uses of “should” and goal-directed uses of “should.” An example of the latter is “If you want the knife to be sharp then you should sharpen it on the whetstone.” But the use of “should” in “In riding a bicycle, when the bicycle leans to the right you should turn the handlebars to the right” is of neither of these varieties. It is perhaps more like the goal-directed kind of “should,” but we are not saying that that is what you should do to achieve the goal of riding a bicycle. Rather, that is part of what is involved *in* riding a bicycle – that is *how* to ride a bicycle. Insofar as we can talk about a goal here at all, it is *defined* by the norms.

Now let us apply this to epistemic norms. We know how to reason. That means that under various circumstances we know what to do in reasoning. This can be described equivalently by saying that we know what we should do. Our epistemic norms are just the norms that describe this procedural knowledge. The way epistemic norms can guide our reasoning without our having to think about them is no longer mysterious. They describe an internalized pattern of behavior that we automatically follow in reasoning, in the same way we automatically follow a pattern in bicycle riding. This is what epistemic norms are. They are the internalized norms that govern our reasoning. Once we realize that they are just one more manifestation of the general phenomenon of automatic behavior governed by internalized norms, epistemic norms should no longer seem puzzling. We would like to have a better understanding of the

psychological process wherein behavior is generated in conformance with internalized norms, and I will say more about this below. But in the meantime, much of the mystery surrounding epistemic norms evaporates once we recognize that the governing process is a general one and its application to epistemic norms and reasoning is not much different from its application to any other kind of action-guiding norms. Of course, unlike most norms our epistemic norms may be innate, in which case there is no process of internalization that is required to make them available for use in guiding our reasoning.⁷

My proposal is that epistemic norms are to be understood in terms of procedural knowledge involving internalized rules for reasoning. This proposal has a close kin in much recent work in psychology and artificial intelligence (AI). Researchers in these fields often model human cognition in terms of *production systems*. These are computational systems described by “condition/action rules,” which tell the system to perform certain actions whenever certain conditions are satisfied.⁸ As I have described them, epistemic norms are condition/action rules and they jointly comprise a production system governing rational belief change.

I have described how our epistemic norms work. This is to describe our *actual* epistemic norms. Internalists typically assume that whatever our actual epistemic norms are, they are the correct epistemic norms. I have taken it to be part of the definition of internalism that our epistemic norms are at least not subject to criticism on externalist grounds. Of course, this is precisely where internalists disagree with norm externalists. Let us turn then to a reconsideration of externalism in the light of our new understanding of epistemic norms.

The Refutation of Externalism

Belief externalism

Now that we understand how epistemic norms work in guiding our reasoning, it is easy to see that they must be internalist norms. This is because when we learn how to do something we acquire a set of norms for doing it and these norms are internalized in a way enabling our central nervous system to follow them in an automatic way without our having to think about them. This has

implications for the content of our norms. For example, I have been describing one of our bike-riding norms as telling us that if the bicycle leans to the right then we should turn the handlebars to the right, but that is not really what we learn when we learn to ride a bicycle. The automatic processing systems in our brain do not have access to whether the bicycle is leaning to the right. What they do have access to are things like (1) our *thinking* that the bicycle is leaning to the right, and (2) certain balance sensations emanating from our inner ear. What we learn is (roughly) to turn the handlebars to the right if we either experience those balance sensations or think on some other basis that the bicycle is leaning to the right. In general, the circumstance-types to which our norms appeal in telling us to do something in circumstances of those types must be directly accessible to our automatic processing systems. The sense in which they must be directly accessible is that our automatic processing system must be able to access them without our first having to make a *judgment* about whether we are in circumstances of that type. We must have non-epistemic access.⁹

This general observation about action-guiding norms has immediate implications for the nature of our epistemic norms. It implies that reason-guiding epistemic norms cannot appeal to external considerations of reliability. This is because such norms could not be internalized. Like *leaning to the right*, considerations of reliability are not directly accessible to our automatic processing systems. There is in principle no way that we can learn to make inferences of various kinds only if they are *in fact* reliable. Of course, we could learn to make certain inferences only if we *think* they are reliable, but that would be an internalist norm appealing to *thoughts* about reliability rather than an externalist norm appealing to reliability itself.¹⁰ Similar observations apply to any externalist norms. Consequently, it is in principle impossible for us to actually employ externalist norms. I take this to be a conclusive refutation of belief externalism.

I introduced the internalism/externalism distinction by saying that internalist theories make justifiedness a function exclusively of the believer's internal states, where internal states are those that are "directly accessible" to the believer. The notion of direct accessibility was purposely left vague, but it can now be clarified. I propose to define internal states as those states that are directly accessible to the mechanisms in our cen-

tral nervous system that direct our reasoning. The sense in which they are *directly* accessible is that access to them does not require us first to have beliefs about them. This definition makes the internalist/externalist distinction precise in a way that agrees at least approximately with the way it has generally been used, although it is impossible to make it agree with everything everyone has said about it because philosophers have drawn the distinction in different ways.

The epistemic norms endorsed by an internalist theory must appeal only to properties of and relations between internal states of the believer. This is not yet enough to characterize internalist norms, however, because an externalist theory might also appeal only to properties of and relations between internalist states of the believer. For instance, probabilism appeals only to the probability of the beliefs held by the believer, and the probability of a belief is a property of it. Internalist theories make the justifiability of a belief a function of the internal states of the believer, in the sense that if we vary anything but his internal states then the justifiability of the belief does not vary. Thus the only properties of and relations between internal states to which internalist norms can appeal are those that cannot be varied without varying the internal states themselves. In other words, they are logical properties of and logical relations between internal states. For instance, if S_1 is the state of believing (P & Q) and S_2 is the state of believing P, then S_1 and S_2 are logically related by the fact that being in S_1 involves believing a conjunction whose first conjunct is believed if one is in state S_2 . Thus we can characterize internalist theories as those proposing epistemic norms that appeal only to logical properties of and logical relations between internal states of the believer.

I have characterized internalist theories in terms of direct accessibility, but I have not said anything in a general way about which properties and relations are directly accessible. It seems clear that directly accessible properties must be in some sense "psychological," but I doubt that we can say much more than that from the comfort of our armchairs. What properties are directly accessible is an empirical question to be answered by psychologists. Despite the fact that we do not have a general characterization of direct accessibility, it is perfectly clear in many specific cases that particular properties to which philosophers have appealed are not directly accessible. In light of this, the preceding refutation of belief externalism can be

applied to a remarkably broad spectrum of theories, and it seems to me to constitute an absolutely conclusive refutation of those theories. I have indicated how it applies to theories formulating epistemic norms in terms of reliability. It applies in the same way to probabilist theories. For example, we saw that many probabilists endorse the *simple rule*:

A belief is epistemically permissible if and only if what is believed is sufficiently probable.

If the simple rule is to provide us with a reason-guiding norm then the probability of a belief must be a directly accessible property of it. No objective probability can have that property. Thus it is impossible to use the simple rule, interpreted in terms of objective probabilities, as a reason-guiding norm. This objection could be circumvented by replacing the simple rule with its “doxastic counterpart”:

A belief is epistemically permissible if and only if the epistemic agent believes it to be highly probable.

But this rule formulates an internalist norm (albeit, an implausible one).¹¹

It might be supposed that we could breathe life back into the simple rule by interpreting it in terms of subjective probability. Here we must be careful to distinguish between subjective probability as actual degree of belief and subjective probability as rational degree of belief. Interpreted in terms of actual degrees of belief, the simple rule would amount to the claim that a belief is justified if and only if it is firmly held, which is an internalist norm, but a preposterous one. Interpreted in terms of rational degrees of belief it becomes an externalist norm. Rational degree of belief is the unique degree of belief one rationally ought to have in a proposition given one’s overall doxastic state, and this is to be understood in terms of prudentially rational betting behavior. As I have indicated, I have serious doubts about the intelligibility of this notion. But even if we waive this objection, ascertaining what this unique rational degree of belief should be is immensely difficult. It seems extremely unlikely that the rational degree of belief one ought to have in a proposition is a directly accessible property of it. If it is not then this version of the simple rule also succumbs to our general objection to belief externalism.

Many other epistemological theories succumb to this objection to belief externalism. For example, Keith Lehrer’s coherence theory is an internalist theory, but an externalist theory can be modeled on it. According to this externalist theory, a person is justified in believing a proposition if and only if that proposition is more probable than each proposition competing with it. But a proposition’s being more probable than any of its competitors is not a directly accessible property of it, and hence the objective version of Lehrer’s theory becomes incapable of supplying us with a reason-guiding norm.

These considerations are efficient in dispatching a wide variety of epistemological theories. All belief-externalist theories succumb to this objection, and a surprising number of internalist theories succumb to it as well. Recall that an internalist theory is any theory proposing epistemic norms that appeal only to logical properties of and logical relations between internal states of the believer. We can make a distinction between properties and relations that are directly accessible and those that are not. A directly accessible property or relation is one to which our automatic processing system has access without our having beliefs about what things have the property or stand in the relation. Not all logical properties or relations are directly accessible, and hence not all internalist theories propose epistemic norms that are internalizable. For instance, a holistic coherence theory adopts a holistic view of reasons according to which a belief is licensed if it is suitably related to the set of *all* the beliefs one holds. A holistic coherence theory requires a relationship between a justified belief and the set of all the beliefs one holds, but that will not normally be a directly accessible property of the justified belief, and hence although the norm proposed by the holistic theory will be an internalist norm, it will not be internalizable. Thus it cannot be reason-guiding.

The general point emerging from all this is that there is a remarkably wide range of epistemological theories succumbing to the simple objection that non-internalist epistemic norms cannot be internalized in the way required in order for them to be reason-guiding. Accordingly, they cannot serve as epistemic norms. No non-internalist theory can provide us with epistemic norms that we could actually use. Correct epistemic norms must be internalist. On the other hand, we have also seen that they must appeal to more than the cognizer’s doxastic state. They must also appeal to his

perceptual and memory states. Thus the correct epistemological theory must endorse some kind of nondoxastic internalist norms.

The endorsement of nondoxastic norms amounts to the rejection of the doxastic assumption, but that has often seemed puzzling. How is it possible for nondoxastic states to justify beliefs when we are not aware that we are in them? We are now in a position to understand how nondoxastic norms are possible. They only seem puzzling because we are implicitly assuming the intellectualist model of the way epistemic norms regulate belief. Given the way epistemic norms actually operate, all that is required is that the input states be directly accessible. Belief states are directly accessible, but so are a variety of nondoxastic states like perceptual states and memory states. Thus there is no reason why epistemic norms cannot appeal to those states, and the rejection of the doxastic assumption and the move to direct realism ceases to be puzzling.

Is there any way to salvage belief externalism in the face of the objection that it cannot give reasonable accounts of first-person reason-guiding epistemic norms? The possibility remains that belief externalism might provide norms for third-person evaluations. I think it is noteworthy in this connection that externalists tend to take a third-person point of view in discussing epistemology. If externalist norms played a role in third-person evaluations, we would then have both externalist and internalist norms that could be applied to individual beliefs and they might conflict. What would this show? It would not show anything – they would just be different norms evaluating the same belief from different points of view. I can imagine a persistent externalist insisting, “Well, if the two sets of norms conflict, which way should we reason – which set of norms should we follow?” But that question does not make any sense. Asking what we should do is asking for a normative judgment, and before we can answer the question we must inquire to what norms the “should” is appealing. To make this clearer consider an analogous case. We can evaluate beliefs from both an epistemic point of view and a prudential point of view. For example, Helen has good reason for believing that her father is Jack the Ripper. Suppose that if she believed that, it would be psychologically crushing. Then we might say that, epistemically, she should believe it, but prudentially she should not. If one then insists upon asking, “Well, should she believe it or not?” the

proper response is, “In what sense of ‘should’ – epistemic or prudential?” Similarly, if externalist and internalist norms conflict and one asks, “Which way should we reason?” the proper response is to ask to which set of norms the “should” is appealing. The point is that different norms serve different purposes, and when they conflict that does not show that there is something wrong with one of the sets of norms – it just shows that the different norms are doing different jobs. The job of internalist norms is reason-guiding, and as such they are the norms traditionally sought in epistemology. Externalist norms (if any sense can be made of them) may also have a point, but they cannot be used to solve traditional epistemological problems pertaining to epistemic justification.

Norm externalism

Recall that there are two kinds of externalism. Belief externalism advocates the adoption of externalist norms. I regard belief externalism as having been decisively refuted by the preceding considerations. Norm externalism, on the other hand, acknowledges that we must employ internalist norms in our reasoning, but proposes that alternative sets of internalist norms should be evaluated in terms of external considerations. For example, it may be alleged that one set of internalist norms is better than another if the first is more reliable in producing true beliefs. Both internalism and norm externalism endorse internalist norms, but they differ in that the internalist alleges that our epistemic norms are not subject to criticism on externalist grounds. It is hard to see how they could be subject to criticism on internalist grounds, so the internalist has typically assumed that our epistemic norms are immune from criticism – whatever our actual epistemic norms are, they are the correct epistemic norms. That, however, seems odd. On the surface, it seems it must be at least logically possible for two people to employ different epistemic norms. They could then hold the same belief under the same circumstances and on the basis of the same evidence and yet the first could be conforming to his norms and the second not conforming to his. If a person’s epistemic norms are always beyond criticism, it would follow that the first person is justified in his beliefs and the second is not, despite the fact that their beliefs are based upon the same

evidence. That would at least be peculiar. Because it seems that it must be possible for different people to employ different epistemic norms, this makes a strong *prima facie* case for norm externalism.

Action-guiding norms are not generally immune from criticism. Typically, action-guiding norms tell us how to do one thing *by* doing something else.¹² For example, knowing how to ride a bicycle consists of knowing what more basic actions to perform – leg movements, arm movements, and the like – by doing which we ride the bicycle. An action that is performed by doing something else is a *nonbasic* action. Norms describing how to perform nonbasic actions can be subject to external evaluation. There may be more than one way to perform the nonbasic action, and some ways may be better (more efficient, more reliable, and so on) than others. If I know how to do it in one way and you know how to do it in another way, you know how to do it better than I if the norms governing your behavior are better than the norms governing mine. For example, we may both know how to hit the target with a bow and arrow, but you may know how to do it more reliably than I.¹³ It thus becomes an empirical question whether acting in accordance with a proposed norm will constitute your doing what you want to be doing and whether another norm might not be better.

Reasoning is not, strictly speaking, an action, but it is something we do, and we do it by doing other simpler things. We reason by adopting new beliefs and rejecting old beliefs under a variety of circumstances. Our norms for reasoning tell us when it is permissible or impermissible to do this. It seems that the norms we actually employ should be subject to external criticism just like any other norms. The norm externalist proposes that we should scrutinize them and possibly replace them by other norms. Because of the direct accessibility problem, we cannot replace them by norms making explicit appeal to reliability, but what we might discover is that (1) under certain circumstances inferences licensed by our natural norms are unreliable, and (2) under certain circumstances inferences not licensed by our natural norms are highly reliable. The norm externalist proposes that we should then alter our epistemic norms, adopting new internalist norms allowing us to make the inferences described under (2) and prohibiting those described under (1).

We must distinguish between two construals of the norm externalist's proposal. He might be tell-

ing us that when we *discover* old reasoning patterns to be unreliable or new reasoning patterns to be reliable then we should alter our norms and our reasoning accordingly. Alternatively, he might be telling us that if old patterns simply *are* unreliable and new patterns *are* reliable, independently of our knowing or believing that they are, then we should alter our reasoning. The first construal seems like an eminently reasonable proposal, and it is one that has been made explicitly by various externalists. For example, in discussing how reliabilist considerations bear on reasoning, Goldman (1981) writes:

At the start a creature forms beliefs from automatic, preprogrammed doxastic processes Once the creature distinguishes between more and less reliable belief-forming processes, it has taken the first step toward doxastic appraisal. . . . The creature can also begin doxastic self-criticism, in which it proposes *regulative* principles to itself. (p. 47)

But this involves a fundamental misconception. Our epistemic norms are not subject to criticism in this way. Particular instances of reasoning are subject to such criticism, and the criticism can dictate changes in that reasoning, but this does not lead to changes in our epistemic norms. This is because unlike other norms, our epistemic norms already accommodate criticism based on reliability. The point is twofold. First, discovering that certain kinds of inferences are unreliable under certain circumstances constitutes a defeater for those inferences and hence makes us unjustified in reasoning in that way, and this is entirely in accordance with our natural unmodified epistemic norms. For example, we discover that color vision is unreliable in dim lighting, and once we discover this we should cease to judge colors on that basis under those circumstances. But this does not require an alteration of our epistemic norms, because color vision only provides us with defeasible reasons for color judgments, and our discovery of unreliability constitutes a defeater for those reasons. This is entirely in accordance with the norms we already have. Second, discovering that some new inferences are reliable under certain circumstances provides us with justification for making those inferences under those circumstances, but this is licensed by the norms we already have. That is precisely what induction is all about. For example, I might discover that I am

clairvoyant and certain kinds of “visions” provide reliable indications of what is about to happen. Once I make this discovery it becomes reasonable for me to base beliefs about the future on such visions. Again, this is entirely in accordance with the norms we already have and does not require us to alter those norms in any way. The general point is that the kinds of reliability considerations to which the norm externalist appeals can lead us to reason differently (refrain from some old inferences and make some new inferences), but this does not lead to any change in our epistemic norms. Epistemic norms are unique in that they involve a kind of built-in feedback having the result that the sort of external criticism that could lead to the modification of other action-guiding norms does not necessitate any modification of epistemic norms.

I have had several externalists respond to this objection by protesting that they do not see the point of distinguishing between considerations of reliability leading us to alter our reasoning and those considerations leading us to alter our norms. But if all the externalist means is that considerations of reliability can lead us to alter our reasoning, then he is not disagreeing with anyone. In particular, he is not disagreeing with paradigmatic internalists like Chisholm and myself. Norm externalism becomes nothing but a pretentious statement of a platitude.

The alternative construal of norm externalism takes it to be telling us that if old patterns of reasoning are unreliable and new patterns are reliable, then regardless of whether we *know* these facts about reliability, we should not reason in accordance with the old patterns and we should reason in accordance with the new patterns. What is the meaning of “should” in this claim? It cannot be taken as a recommendation about how to reason, because it is not a recommendation anyone could follow. We can only alter our reasoning in response to facts about reliability if we are apprised of those facts. However, normative judgments do not always have the force of recommendations. That is, they are not always intended to be action-guiding. This is connected with the distinction that is often made in ethics between subjective and objective senses of “should.” To say that a person subjectively should do X is to say, roughly, that given what he believes (perhaps falsely) to be the case he has an obligation to do X. To say that he objectively should do X is to say, roughly, that if he were apprised of all the relevant facts then he

would have an obligation to do X. Judgments about what a person subjectively should do can serve as recommendations, but judgments about what a person objectively should do can only serve as external evaluations having some purpose other than guiding behavior.¹⁴ The subjective/objective distinction can be regarded as a distinction between evaluating the person and evaluating his act. The subjective sense of “should” has to do with moral responsibility, while the objective sense has to do with what act might best have been performed.

We can draw a similar subjective/objective distinction in epistemology. The epistemic analogue of moral responsibility is epistemic justification. A person is being “epistemically responsible” just in case his beliefs are justified. In other words, epistemic justification corresponds to *subjective* moral obligation. What determines whether a belief is justified is what else the epistemic agent *believes* about the world (and what other directly accessible states he is in) – not what is in fact true about the world. This seems to show that whatever considerations of *de facto* reliability may bear upon, it is not epistemic justification. They must instead bear upon the epistemic analogue of objective obligation. What is that analogue? There is one clear analogue – objective epistemic justification is a matter of what you should believe if you were apprised of all the relevant truths. But what you should believe if you were apprised of all the relevant truths is just *all the truths*. In other words, the epistemic analogue of objective justification is *truth*. There is nothing here to give solace to a norm externalist.

Goldman (1981) draws a somewhat different distinction between two senses of “justified” in epistemology. He distinguishes between “theoretical” evaluations of reasoning and “regulative” evaluations (the latter being reason-guiding). He suggests that the theoretical sense of justification is the sense required for knowledge and that it is to be distinguished from the reason-guiding sense. He suggests further that his reliabilist theory concerns the theoretical sense. The proposal is that it is knowledge that provides the point of a norm externalist’s evaluation of epistemic norms in terms of considerations of reliability unknown to the epistemic agent. I do not believe that, but even if it were true it would not affect my overall point. The sense of epistemic justification with which I am concerned in this book is the reason-guiding sense, and if it is acknowledged that norm

externalism bears only upon another sense of justification then my main point has been conceded.

Epistemological relativism and the individuation of concepts

The apparent failure of norm externalism leaves us with a puzzling problem. Internalists have typically assumed that whatever epistemic norms we actually employ are automatically correct. But that seems hard to reconcile with the seemingly obvious fact that it is at least logically possible for different people to employ different norms. Surely, if Smith and Jones believe *P* for the same reasons, they are either both justified or both unjustified. There is no room for their justification to be relative to idiosyncratic features of their psychology resulting in their employing different epistemic norms. This seems to imply that there is just one set of correct epistemic norms, and the norms a person actually employs may fail to be correct. This conclusion would seem to be obvious if it were not for the fact that there is no apparent basis for criticizing a person's norms. That is precisely what norm externalism tries unsuccessfully to do. The reliabilist considerations to which the norm externalist appeals are the only plausible candidates for considerations of use in criticizing and correcting epistemic norms, and we have seen that our epistemic norms cannot be corrected in this way. Of course, I might criticize Jones' norms simply because they disagree with mine, but he could equally criticize mine because they disagree with his. Are we committed to a thoroughgoing epistemological relativism then? That is at least unpalatable.

The solution to the problem of relativism can be found by turning to a different problem. This is the problem of how concepts are individuated. The standard view takes concepts to be individuated by their truth conditions. The claim of this theory is that what makes a concept the concept that it is are the conditions that must be satisfied for something to exemplify that concept. These conditions comprise its truth conditions. The precise content of the truth condition theory of concepts deserves closer inspection than it usually receives. There is one sense in which the truth condition theory of concepts is correct but also completely trivial and uninteresting. The truth condition of the concept *red* is the condition of *being red*, and the truth condition of the concept

blue is the condition of *being blue*. The following is undeniable:

red = blue if and only if *being red = being blue*

but it is hardly illuminating. Rather than explaining the concepts, the truth conditions presuppose the concepts. We might just as well define the "identity condition" of a physical object to be the condition of *being that object* and then claim that physical objects are individuated by their identity conditions. That is about as unilluminating as a theory can be.

Typically, philosophical logicians slide back and forth between the vacuous claim that concepts are individuated by their truth conditions and the considerably more contentious claim that concepts can be informatively characterized by (and only by) giving truth condition analyses of them. A truth condition analysis of a concept is an informative statement of necessary and sufficient conditions for something to exemplify the concept. I think it is fair to say that many philosophical logicians do not clearly distinguish between the vacuous claim and the contentious claim, or at least take the vacuous claim to somehow directly support the contentious claim. But I see no reason to think there is any connection between the two claims.

There is another strand to this story. Traditionally, the only logical relations between concepts that were recognized by philosophers were entailment relations. Concepts, as "logical items," were supposed to be individuated by their logical properties, and it seemed that the only logical properties concepts possessed were those definable in terms of their entailment relations to other concepts. This generates the picture of a "logical space" of concepts, the identity of a concept being determined by its position in the space, and the latter being determined by its entailment relations to other concepts. The claim that concepts must have definitions is just a more specific version of this general picture – one alleging that the position of a concept in logical space is determined not just by one-way entailments but by two-way logical equivalences. Some version of this picture has been prevalent throughout much of twentieth century philosophy, and it still plays a prominent role in philosophical logic. I will call this general picture of the individuation of concepts *the logical theory of concepts*. It has typically been either confused with or identified with the truth condition theory.

The simplest objection to all of this is that most concepts do not have the kind of definitions required by the logical theory of concepts. Analytic philosophy in the mid-twentieth century concerned itself almost exclusively with the search for such definitions, and if we can learn anything from that period it is that the search was largely in vain. It is a very rare concept that can be given an informative definition stating truth conditions. The importance of this simple objection cannot be overemphasized. Most concepts do not have definitions. For reasons I find mysterious, many philosophers seem to just ignore this and go on pretending that the logical theory of concepts is correct.

We can also raise a more purely epistemological problem for the logical theory of concepts, and I will now spend some time developing this problem. In general, the logical theory cannot make sense of reasons. To see this, let us begin with *prima facie* reasons. The logical theory appears to lead directly to the impossibility of *prima facie* reasons. I assume that what makes something a good reason for holding a belief is a function of the content of the belief. If the content of the belief is determined by entailment relations, then those entailment relations must also determine what are good reasons for holding that belief. The only kinds of reasons that can be derived from entailment relations are reasons that are themselves entailments – conclusive reasons. Thus we are forced to the conclusion that all reasons must be entailments. But this must be wrong, because we have seen that we cannot solve epistemological problems in terms of conclusive reasons. Justified belief makes essential appeal to defeasible reasoning.

We might try distinguishing between “formal reasons” that derive from principles of logic and apply equally to all concepts, and “substantive reasons” that are specific to individual concepts and reflect the contents of those concepts. The preceding argument is really only an argument that the logical theory of concepts is incompatible with there being nonconclusive substantive reasons. Thus we could render the logical theory of concepts compatible with defeasible reasoning if it could be maintained that all legitimate defeasible reasons are formal reasons. The only plausible way of defending this claim is to maintain that the only legitimate defeasible reasons are inductive reasons and to insist that inductive reasons are formal reasons. This is to take induction to be a species of logic. On this view, there are two kinds of logic

– deductive and inductive – and each generates formal reasons that pertain to all concepts and hence need not be derivable from the contents of individual concepts. For example, a conjunction (P & Q) gives us a reason for believing its first conjunct P regardless of what P and Q are. Similarly, it was traditionally supposed that inductive reasons are formal reasons pertaining equally to all concepts. This absolves us from having to derive inductive *prima facie* reasons from the essential properties of the concepts to which the reasons apply.

Unfortunately, this attempt to render the logical theory of concepts compatible with induction fails. I have pointed out elsewhere that induction does not apply equally to all concepts. Inductive reasoning must be restricted to projectible concepts. There is no generally accepted theory of projectibility, but it is generally recognized that what makes a concept projectible is not in any sense a “formal” feature of it. The simplest argument for this was given long ago by Nelson Goodman (1955). Define:

x is *grue* if and only if either (1) x is green and first examined before the year 2000, or (2) x is blue and not first examined before the year 2000.

x is *bleen* if and only if either (1) x is blue and first examined before the year 2000, or (2) x is green and not first examined before the year 2000.

“Grue” and “bleen” are not projectible. For example, if we now (prior to the year 2000) examine lots of emeralds and find that they are all green, that gives us an inductive reason for thinking that all emeralds are green. Our sample of green emeralds is also a sample of *grue* emeralds, so if “grue” were projectible then our observations would also give us a reason for thinking that all emeralds are *grue*. These two conclusions together would entail the absurd consequence that there will be no emeralds first examined after the year 2000. It follows that “grue” is not projectible. Now the thing to notice is that “blue” and “green” are definable in terms of “grue” and “bleen” in the precisely the same way “grue” and “bleen” were defined in terms of “blue” and “green”:

x is green if and only if either (1) x is *grue* and first examined before the year 2000, or (2) x is

bleen and not first examined before the year 2000.

x is blue if and only if either (1) x is bleen and first examined before the year 2000, or (2) x is grue and not first examined before the year 2000.

Thus the *formal* relationships between the pair “blue,” “green” and the pair “grue,” “bleen” are symmetrical, and hence we cannot distinguish the projectible from the nonprojectible by appealing only to formal properties of the concepts. Projectibility seems to have essentially to do with the content of the concepts. Therefore, any explanation for the existence of inductive prima facie reasons must make reference to the particular concepts to which the reasons apply, and hence, on the logical theory of concepts, inductive prima facie reasons become as mysterious as any other prima facie reasons.

There is of course the further point, defended earlier, that epistemology requires more prima facie reasons than just inductive ones. Thus even if inductive reasons had turned out to be formal reasons, that would not entirely solve the problem of the possibility of prima facie reasons.

The next thing to notice is that the logical theory of concepts makes conclusive reasons just as mysterious as prima facie reasons. This has generally been overlooked, but it is really rather obvious. Epistemologists have noted repeatedly that logical entailments do not always constitute reasons. Some entailments are conclusive reasons and others are not reasons at all. The latter is because P may entail Q without the connection between P and Q being at all obvious. For example, mathematicians have proven that the Axiom of Choice entails Zorn’s Lemma. These are abstruse mathematical principles apparently dealing with quite different subject matters, and just looking at them one would not expect there to be any connection between them. If, without knowing about the entailment, one were so perverse as to believe Zorn’s Lemma on the basis of the Axiom of Choice, one would not be justified in this belief. Once the entailment is known, you can become justified in believing Zorn’s Lemma *partly* by appeal to the Axiom of Choice, but your full reason for believing Zorn’s Lemma will be the conjunction of the Axiom of Choice and the proposition that if the Axiom of Choice is true then Zorn’s Lemma is true. You are believing

Zorn’s Lemma on the basis of this conjunction rather than just on the basis of the Axiom of Choice. You can never become justified in believing Zorn’s Lemma on the basis of the Axiom of Choice alone, so the latter is not a reason for the former.

On the other hand, some entailments do provide reasons. If I justifiably believe both P and $(P \supset Q)$, I *can* justifiably believe Q on the basis of these other two beliefs. In this case I do not have to believe Q on the basis of the more complicated belief:

P and $(P \supset Q)$ and if $[P \ \& \ (P \supset Q)]$ then Q .

To suppose that each instance of reasoning in accordance with *modus ponens* must be reconstructed in this way would lead to an infinite regress.¹⁵ Thus some entailments are conclusive reasons and others are not. But the logical theory of concepts gives us no way to make this distinction. It characterizes concepts in terms of their entailment relations to other concepts, but, *a fortiori*, all entailment relations are entailment relations. There is nothing about the entailment relations themselves that could make some of them reasons and others not. Thus conclusive reasons become just as mysterious as prima facie reasons on the logical theory of concepts. This seems to indicate pretty conclusively that the logical theory of concepts is wrong. There has to be more to concepts than entailment relations.

To argue that the logical theory of concepts is wrong is not yet to say what is right. The theory I want to endorse in its place is *the epistemological theory of concepts*. This theory begins by noting that concepts are both logical and epistemological items. That is, concepts are the categories whose interrelationships are studied by logic, and they are also the categories in terms of which we think of the world. The interrelationships studied by logic can all be reduced to entailment relations. Thus logic need not take note of any other features of concepts. Logic can get along with a cruder picture of concepts than can epistemology. But a complete account of concepts must accommodate both logic and epistemology. There is good reason to think that the role of concepts in epistemology is fundamental. Not all entailment relations are conclusive reasons, but it seems likely that all entailment relations derive from “simple” entailment relations, where the latter are just those that are conclusive reasons. Thus a theory of concepts

adequate for epistemology will very likely be adequate for logic as well. The question then becomes, "What kind of theory of concepts is adequate for epistemology?" In epistemology, the essential role of concepts is their role in reasoning. They are the categories in terms of which we think of the world, and we think of the world by reasoning about it. This suggests that concepts are individuated by their role in reasoning. What makes a concept the concept that it is is the way we can use it in reasoning, and that is described by saying how it enters into various kinds of reasons, both conclusive and *prima facie*. Let us take the *conceptual role* of a concept to consist of (1) the reasons (conclusive or *prima facie*) for thinking that something exemplifies it or exemplifies its negation, and (2) the conclusions we can justifiably draw (conclusively or *prima facie*) from the fact that something exemplifies the concept or exemplifies the negation of the concept. My proposal is that concepts are individuated by their conceptual roles. The essence of a concept is to have the conceptual role that it does. If this is right, the explanation for how there can be such things as *prima facie* reasons becomes trivial. *Prima facie* reasons are primitive constituents of the conceptual roles that characterize concepts. *Prima facie* reasons need not have an origin in something deeper about concepts, because there is nothing deeper. In an important sense, there is nothing to concepts over and above their conceptual role. To describe the conceptual role of a concept is to give an analysis of that concept, although not a truth condition analysis.¹⁶

I think it is undeniable that concepts are individuated by their conceptual roles, and not (at least in any non-vacuous way) by their truth conditions. But some further explanation for all of this is required. *Why* are concepts individuated in this way? I will shortly propose an answer to this question. For the moment, however, I will simply take it as established that concepts are individuated in this way. The importance of this theory of concepts for the matters at hand is that it lays to rest the specter of epistemological relativism. Epistemological relativism is the view that (1) different people could have different epistemic norms that conflict in the sense that they lead to different assessments of the justifiedness of the same belief being held on the same basis, and (2) there is no way to choose between these norms. The epistemological theory of concepts enables us to escape any such relativism. Because concepts are individuated by their conceptual roles, it becomes

impossible for people's epistemic norms to differ in a way that makes them conflict with one another. The epistemic norms a person employs in reasoning determine what concepts he is employing because they describe the conceptual roles of his concepts. If two people reason in accordance with different sets of epistemic norms, all that follows is that they are employing different concepts. Thus it is impossible for two people to employ different epistemic norms in connection with the same concepts. Their conceptual frameworks are determined by their epistemic norms. Epistemological relativism is logically false.¹⁷

Conclusions

To summarize the discussion of externalism, one can be an externalist by being either a belief externalist or a norm externalist. These exhaust the ways in which externalist considerations might be brought to bear on our epistemic norms. The belief externalist tries to formulate epistemic norms directly in terms of externalist considerations, but it is impossible to construct reason-guiding norms in this way. The norm externalist proposes instead to recommend changes in reason-guiding norms on the basis of considerations of reliability. But this appeal to reliability is redundant because it is already accommodated by our unmodified internalist norms. Thus, as far as I can see, externalism has nothing to contribute to the solution to traditional epistemological problems. Justified beliefs are those resulting from normatively correct reasoning. Consequently, any evaluation of the justifiedness of a belief must be reason-guiding and hence must be beyond the pale of externalism.

Man as a Cognitive Machine

I have described how epistemic norms work and how they are related to concepts, but we may be left wondering why all of this should be the case. A fuller understanding of the nature of epistemic norms can be obtained by seeing how they are integrated into the broader picture of man as a cognitive machine. I take it for granted that man is a kind of biological information processor. Considerable light can be thrown on human epistemology by reflecting on the workings of cognitive machines in general.

Suppose we were undertaking the feat of building an “intelligent machine” that could interact with its surroundings, learn from experience, and survive in a reasonably hostile environment. Let’s call our machine “Oscar.” What would we have to put into Oscar to make him work? At the very least we would have to provide him with ways of sensing the environment and thinking about the world. It is worth pursuing some of the details.

Oscar I

We must begin by incorporating sensors much like our sense organs so that Oscar can respond to states of the environment, and we might even call these sensors “sense organs.” We must also incorporate “reasoning” facilities, both deductive and inductive. And we must incorporate some sort of conative structure to provide goals for Oscar to attempt to realize. If Oscar is to survive in a hostile environment, it would also be wise to provide sensors that respond to conditions under which he is in imminent danger of damage or destruction. We might call these “pain sensors.” Oscar could then have built-in “fight or flight” responses elicited by the activation of his pain sensors.

I have described Oscar as thinking about the world. That involves a system of mental representation – what we might call a “language of thought.”¹⁸ For Oscar to have a thought is for him to “entertain” a sentence in his language of thought and treat it in a certain way. Without going into details, we can suppose abstractly that for Oscar to have a thought is for him to have a sentence in his language of thought residing in his “B-box.”¹⁹ Adopting a computer metaphor, we can think of the latter as a memory location. Oscar’s thoughts and beliefs must be causally related to his environment and his behavior. On the one hand, Oscar must be constructed in such a way that the stimulation of his sensory apparatus tends to cause him to acquire certain beliefs. On the other hand, Oscar’s having appropriate beliefs must tend to cause him to behave in corresponding ways. To describe these causal connections will be to describe his facilities for “pure reasoning” (his epistemology) and his facilities for practical reasoning.

Let us call the machine resulting from this stage of design “Oscar I.”²⁰

Oscar II

Oscar I could function reasonably well in a congenial environment. But in an environment that is both reasonably complex and reasonably hostile, Oscar I would be doomed to early destruction. He would be easy meat for wily machinivores. The difficulty is this. To be effective in avoiding damage, Oscar I must not only be able to *respond* to the stimulation of his pain-sensors when that occurs – he must also be able to *predict* when that is apt to occur and avoid getting into such situations. He must be able to exercise “foresight.” As we have constructed him, Oscar I has the ability to form generalizations about his environment as sensed by his sense organs, but he has no way to form generalizations about the circumstances in which his pain-sensors are apt to be activated. This is because Oscar I has no direct way of knowing when his pain-sensors are activated – he has no way of “feeling pain.” As I have described them, the pain-sensors cause behavioral responses directly and do not provide input to Oscar’s cognitive machinery. If Oscar is to be able to avoid pain rather than merely respond to it, he must be able to tell when he is in pain and be able to form generalizations about pain. To do this he needs another kind of sensor – a “pain-sensor sensor” that detects when the pain-sensors are activated. (Of course, the pain-sensors can themselves be pain-sensor sensors if they send their outputs to more than one place. We do not need a separate organ to sense the operation of the first organ.) Suppose we build these into Oscar I, renaming him Oscar II. This gives him a rudimentary kind of self-awareness. If the conative structure of Oscar II is such that he is moved to avoid not only the current activation of his pain-sensors but their anticipated activation as well, then this will enable him to avoid getting into situations that would otherwise result in his early demise.

It is illuminating to note that the difference between Oscar I and Oscar II is roughly the difference between an amoeba and a worm. Amoebas only *respond* to pain (or more conservatively, what we can regard as the activation of their pain-sensors) – worms can learn to avoid it. The learning powers of worms are pretty crude, proceeding entirely by simple forms of conditioning, but we have said nothing about Oscar that requires him to have greater learning powers.

Beginning with Oscar II we can distinguish between two kinds of sensors. First, Oscar II has *external sensors* to sense the world around him. These are of two kinds. He has ordinary perceptual sensors, and he also has pain-sensors that respond to environmental stimuli that tend to indicate impending damage to his body. Oscar II also has an *internal sensor* to sense the operation of his pain-sensors. His internal sensor could be described as a 'higher-order sensor' because it senses the operation of another sensor.

Oscar III

Oscar II is still a pretty dumb brute. I have described him as sensing his physical environment and forming generalizations on that basis. But he does not do a very good job of that. The trouble is that he can only take his perception of the environment at face value. If his "red sensor" provides the input "red" to his cognitive machinery, he can relate that to various generalizations he has formed concerning when there are red things about, and he can also use the input to form new generalizations. But the generalizations at which he will arrive will be crude affairs. He will have no conception of the environment fooling him. For example, he will be unable to distinguish between a machine-eating tiger and a mirror image of a machine-eating tiger. All he will be able to conclude is that some tigers are dangerous and others are not. We, on the other hand, know that all tigers are dangerous, but that sometimes there is no tiger there even though it looks to us like there is. Oscar II has no way of learning things like this. He has no way of discovering, for example, that his red sensor is not totally reliable. This is because, at least until he learns a lot about micromechanics, he has no way to even know that he has a red sensor or to know when that sensor is activated. He responds to the sensor in an automatic way, just as Oscar I responded to his pain-sensors in an automatic way. If Oscar II is to acquire a sophisticated view of his environment, he must be able to sense the activation of his red sensor.²¹ That will enable him to discover inductively that his red sensor is sometimes activated in the absence of red objects.

This point really has to do with computing power. Given sufficient computing power, Oscar might be able to get by, forming all of his generalizations directly on the basis of the output of his external sensors. His generalizations would parallel

the kind of "phenomenalistic generalizations" required by the phenomenalist epistemologies championed in the first half of this century by such philosophers as Rudolf Carnap, Nelson Goodman, and C. I. Lewis.²² The most salient feature of such generalizations would be their extraordinary complexity. Just imagine what it would be like if instead of thinking about physical objects you had to keep track of the world entirely in terms of the way things appear to you and your generalizations about the world had to be formulated entirely in those terms. You could not do it. Human beings do not have the computational capacity required to form and confirm such complex generalizations or to guide their activities in terms of them. Instead, human beings take perceptual input to provide only *prima facie* reasons for conclusions about their physical environment. This allows them to split their generalizations into two parts. On the one hand they have generalizations about the relations between their perceptual inputs and the state of their environment, and on the other hand they have generalizations about regularities within the environment that persist independently of perception of the environment. The advantage of dividing things up in this way is that the two sets of generalizations can be adjusted in parallel to keep each manageably simple under circumstances in which purely phenomenalistic generalizations would be unmanageable. Epistemologically, we begin by trusting our senses and taking their pronouncements to be indicative of the state of the world. More formally, appearance provides us with *prima facie* reasons for judgments about the world and, initially, we have no defeaters for any of those judgments. Making initial judgments in this way we find that certain generalizations are approximately true. If (a) we can make those generalizations exactly true by adjusting some of our initial judgments about the world, and (b) we can do it in such a way that there are simple generalizations describing the circumstances under which things are not as they appear, we take that as a defeater for the initial perceptual judgments that we want to overturn and we embrace the two sets of generalizations (the generalizations about the environment and the generalizations about the circumstances under which perception is reliable). The result is a considerable simplification in the generalizations we accept and in terms of which we guide our activities.²³ A secondary effect is that once we acquire evidence that a generalization is approximately true, there is

a “cognitive push” toward regarding it as exactly true.

The logical form of what goes on here is strikingly similar to traditional accounts of scientific theory formation. On those accounts we begin with a set of data and then we “posit theoretical entities” and construct generalizations about those entities with the objective of constructing a theory that makes correct predictions about new data. There is a formal parallel between this picture and our thought about physical objects. Physical objects play the role of theoretical entities, our sensory input provides the data, and we try to adjust the generalizations about physical objects and the “bridge rules” relating physical objects and sensory input, in such a way that we can make correct predictions about future sensory input. Of course, all of this is to over-intellectualize what goes on in human thought. We do not invent physical objects as theoretical entities designed to explain our sensory inputs. We just naturally think in terms of physical objects, and our conceptual framework makes that epistemologically legitimate independent of any reconstruction of it in terms of scientific theory formation. My point is merely that the logical structure is similar. From an information-processing point of view, the adoption of such a logical structure gives us an additional degree of freedom (the physical objects, or the theoretical entities) that can be adjusted to simplify the associated generalizations and thus minimize the computational complexity of using those generalizations to guide activity.²⁴

The point of all this is that to acquire the kind of manageable generalizations about the environment that will enable him to keep functioning and achieve his built-in goals, an intelligent machine must be able to sense the operation of his own sensors. Only in that way can he treat the input from these sensors as defeasible and form generalizations about their reliability, and the need to treat them this way is dictated by considerations of computational complexity. Let’s build such second-order sensors into Oscar II and rename him Oscar III. He thus acquires a further degree of self-awareness. The difference between Oscar II and Oscar III may be roughly parallel to the difference between a bird and a cat. Kittens quickly learn about mirror images and come to ignore them, but birds will go on attacking their own reflections until they become exhausted.

Although Oscar III has second-order sensors sensing the operation of his first-order “percep-

tual” sensors, this does not mean that he can respond to his perceptual sensors only by sensing their operation – that is the mistake of the foundations theorist. In the ordinary course of events Oscar III can get along fine just responding mechanically to his perceptual sensors. To attend to the output of a sensor is to utilize (in cognition) the output of a higher order sensor that senses the output of the first sensor. Oscar III need not attend to the output of his perceptual sensors under most circumstances because doing so would not alter his behavior (except to slow him down and make him less efficient). He need attend only to the output of his second-order sensors under circumstances in which he has already discovered that his first-order sensors are sometimes unreliable.²⁵ This is related to the fact that Oscar III will automatically have reason to believe that sense perception is generally reliable. (Of course, he might be wrong about this.)

The cognitive role of the pain-sensors is a bit different from that of the perceptual organs. Oscar III will function best if he almost always attends to the output of his pain-sensors. These play a different kind of role than the perceptual sensors. Their role is not just one of fine-tuning. Except in “emergency situations” in which all cognitive powers are brought to bear to avoid a permanent systems crash, Oscar III should always be on the lookout for new generalizations about pain, and this requires that he almost always be aware of when his pain-sensors are activated. This parallels the fact that in human beings we are much more aware of our pains than of our visual sensations. We generally “look through” our visual sensations at the world and do not think about the sensations themselves.

I have attributed two kinds of self-awareness to Oscar III – he has the ability to sense the activation of his pain-sensors and also to sense the activation of his perceptual organs. These proceed via “internal” or “higher order” sensors. The important thing to realize is that there are simple explanations for why such self-awareness will make an intelligent machine work better. Other kinds of self-awareness may also be either desirable or necessary. I have not described Oscar III as having any awareness of what goes on internally after he acquires perceptual input or pain stimuli. In particular, I have not described him as having any way of sensing the operation of those cognitive processes whereby he forms generalizations on the basis of his perceptual inputs and pain stimuli.

But such awareness seems to be required for two reasons. First, consider defeasible reasoning. In defeasible reasoning we reason to a conclusion, and then subsequent reasoning may lead to new conclusions that undercut the original reasoning and cause us to retract the original conclusion. In order for such negative feedback to work, the cognitive agent must be able to sense and keep track of his reasoning processes. Actually, humans are not terribly good at this. We forget our reasons rather rapidly, and we often fail to make appropriate corrections even when we remember our reasons.²⁶ We would probably work better if we could keep better track of our reasoning processes. At any rate, the general point seems clear. The ability to sense his own reasoning processes will be required in order for Oscar to indulge in defeasible reasoning, and defeasible reasoning seems to be required by any kind of sophisticated epistemology.

There is another reason a well-functioning cognitive agent must be able to sense his own reasoning processes to some extent. A cognitive agent does not try to gather information at random – he normally seeks to answer specific questions (motivated ultimately by conative considerations). Built-in epistemic norms will determine what kind of reasoning he is *permitted* to pursue, but they do not dictate what reasoning he *must* pursue. Effective problem solving (at least in humans) involves the use of reasoning strategies rather than random permissible reasoning, and we acquire such strategies by learning about how best to search for solutions to various kinds of problems. To learn such things we must be aware of how we proceed in particular cases so that we can make generalizations about the efficacy of the search procedures employed.

Mental representations

I described the Oscarites as having a language of thought to encode information. This language of thought is a representational system. It must, among other things, provide ways of thinking about particular objects and ascribing properties to them. I will call ways of thinking of objects *mental representations*. Philosophy has traditionally adopted a rather parochial view of mental representations, often recognizing only one way of thinking of an object – as the unique object having a certain combination of properties. This is to think

of an object “under a description.” In fact, this is one of the least common ways of thinking of objects. This is most easily seen by considering how often you can actually find a description that uniquely picks out an object without the description itself involving a way of thinking about another object. For instance, I can think of my mother as “the mother of me,” but that only works insofar as I already have some way of thinking of myself. It is often fairly easy to propound descriptions that pick out objects uniquely as long as we are allowed to build into the descriptions relations to other objects. But to get such descriptions going in the first place, we must begin with some other ways of thinking of at least some objects. Those other ways could involve descriptions provided those descriptions did not make reference to further objects, but I challenge the reader to find even one such description. If we cannot find such descriptions, then it seems clear that they do not constitute the mental representations in terms of which we think of objects.

The unavoidable lesson to be learned from the paucity of descriptions is that we have nondescriptive ways of thinking of at least some objects. We do not have to look very far to find some of those nondescriptive mental representations. First, consider perception. When I see an object and make a judgment about it, I do not usually think of that object under a description – not even a description like “the object I am seeing” (I am typically seeing many different objects at any one time). Instead, I just focus my attention on the object and have a thought whose closest expression in English is something like “That is a table.” In a case like this, my visual experience involves what we might call a “percept” of an object, and I think of the object in terms of that percept. That percept is my mental representation of the perceived object, and it is a constituent of my thought.²⁷ Percepts are not descriptions, so this is an example of a nondescriptive mental representation.

A percept can only represent an object while that object is being perceived. If I later see another object that looks precisely the same way to me, then precisely the same percept will recur, but this time it will represent the new object. I can, however, continue to think about the original object after I am no longer perceiving it. When I do that I am no longer thinking of it in terms of the percept, so I must be employing a different kind of mental

representation. This new mental representation still need not be a description. Once I have become able to think about an object in some way or other, I can continue to think about it even when that original mental representation is no longer available to me. This is clear in the case of objects originally represented by percepts, but it is equally true of objects originally thought about under descriptions. Thinking of an object under a description, I may acquire a wide variety of beliefs about it, but I may eventually forget the original description. For instance, I might have first come to think of Christopher Columbus under some description like "The man my teacher is talking about," but I can no longer remember just what description I might have used and I may no longer remember that Christopher Columbus satisfies that description. Such forgetfulness does not deprive me of my ability to think of Christopher Columbus. I have a nondescriptive way of thinking of Christopher Columbus. Such nondescriptive ways of thinking of an object are parasitic on originally having some other way of thinking of the object (either perceptual or descriptive), but they are distinct from those other ways. I call these nondescriptive ways of thinking of objects "*de re* representations," and I have written about them at length elsewhere.²⁸

The above remarks are largely remarks about the phenomenology of human thought. They amount to the observation that we rarely think of objects under descriptions, but we do have both perceptual and nonperceptual nondescriptive ways of thinking of objects. As a remark about human psychology, this seems obviously correct, but it may also seem philosophically puzzling. How can there be such nondescriptive ways of thinking of objects? I think, however, we can make the puzzlement go away by reflecting on Oscar. Mental representations are just singular terms in the language of thought. If it can be shown that there is no obstacle to constructing Oscar in such a way that his language of thought contains such singular terms, then there should be no reason to be suspicious of the claim that human beings employ such mental representations. From an information-processing point of view we can think of *de re* representations as pigeon holes (memory locations) into which we stuff properties as we acquire reasons to believe that the objects represented have those properties. Properties may drop out of the pigeon holes if they are not used occasionally (i.e., we forget). In order to establish a pigeon hole as

representing a particular object we must begin by thinking of the object in some other way, and that initial way of thinking of the object will be the first thing to be put into the pigeon hole. For example, we might begin by thinking of an object under a description, from that acquire a *de re* representation of the object, and then we might eventually forget the original description and only be able to think of the object under the *de re* representation. From an information-processing point of view there are good reasons (having to do with the efficient use of memory) for incorporating something like *de re* representations into the language of thought used by an intelligent machine, and we would be well advised to equip Oscar with some such device.

Another important representational device is that involved in first-person beliefs. Numerous philosophers have observed that, although we can think of ourselves in more familiar ways (e.g., under descriptions, or perceptually), we can also have thoughts about ourselves in which we do not think of ourselves in any of those ways.²⁹ I will follow David Lewis in calling these *de se* beliefs.³⁰ The mental representations employed in *de se* beliefs will be called *de se* representations. The existence of *de se* representations is illustrated by the following example (due to John Perry (1979)):

I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally, it dawned on me. I was the shopper I was trying to catch.

What happened when Perry realized that he was the shopper with the torn sack? He came to believe an identity, viz., that the shopper with the torn sack was the same person as he himself. This identity involves thinking of the same individual (himself) in two different ways (with two different mental representations) and believing that they *are* the same individual. The first mental representation is a straightforward descriptive representation – Perry thinks of himself as "the shopper with the torn sack." But the second representation, in which he thinks of himself as "me, myself," is a unique representation different from the kinds of

representations we can employ in thinking of other things. I can think of myself under a description, or perceptually (e.g., I may see myself in a mirror), or in terms of a *de re* representation, but whenever I think of myself in one of those ways I may fail to realize that it is myself that I am thinking of. To realize that I am thinking of myself is to relate the mental representation I am employing to a special way of thinking of myself – my *de se* representation of myself.

Why do I have such a special way of thinking of myself? *De se* representations are essential elements of the language of thought of any sophisticated cognizer (human or otherwise). This is most easily illustrated by considering the conative aspects of Oscar. The purpose of providing Oscar with cognitive powers is to enable him to achieve built-in goals. This is accomplished by combining his beliefs with a conative structure consisting of preferences, desires, aversions, and so on, and natural inclinations to behave in specifiable ways in the presence of particular combinations of beliefs and conations. The problem now is to construct appropriate behavioral tendencies of the latter sort in a creature lacking *de se* beliefs. I claim that it cannot be done. Practical reasoning consists of forming the intention to do something under specified circumstances, forming the belief that you are in such circumstances, and then performing the action. We must supply the cognitive agent with rules for the formation of appropriate intentions and beliefs. The intentions are *conditional* intentions to do something *if* some particular condition is satisfied. To be useful, the conditions must concern the agent's situation and not just the general state of the universe. They must involve the *agent's* being in the specified circumstances. The rules for the formation of such conditional intentions must be constructed in such a way that the condition involves a mental representation of the agent. The representation cannot be one that just happens to represent the agent, because the rules must be constructed prior to any contingent knowledge of the world. For instance, the rules for intention formation cannot proceed in terms of definite descriptions that just happen to represent the agent, because it cannot be predicated ahead of time which of these will represent the agent. The rules themselves must require that the intentions involve a mental representation of the agent, and at the time the rules are being constructed it cannot be predicted what mental representations will represent the agent as a matter of contingent fact,

so the rules for intention formation must employ a mental representation that represents the agent necessarily. That is precisely what *de se* representations are.

It is illuminating to illustrate this by considering chess-playing programs. Sophisticated chess-playing programs learn and get better as they play. Existing programs do not involve any kind of *de se* representation, and that may seem to be a counterexample to the claims just made. But the reason such programs need not involve *de se* representation is that, in the appropriate sense, everything in their vocabulary describes the situation of the chess-playing computer, and so there is no distinction to be drawn between its situation and the general state of the universe. Contrast this with a computer running a more sophisticated kind of chess-playing program that learns not just by playing but also by witnessing games played by other computers. In this case the computer must be able to distinguish between its own game and games it is merely witnessing. Its own game must somehow be tagged *as* its own. In effect, this involves *de se* representation.

My conclusion is that practical reasoning requires an agent to form beliefs about his *own* current situation and form intentions about what to do if his current situation is of a particular sort. It does no good to have general existential beliefs about the state of the universe if those beliefs cannot be related to the agent, and it is precisely the latter that cannot be done without *de se* beliefs. *De se* representations are required in order to make conation and intellection mesh properly. Thus *de se* beliefs are essential in the construction of a sophisticated cognitive/conative machine. I take it that that is why they play a central role in human thought.

It is somewhat illuminating to consider just what we have to do to equip Oscar with *de se* thought. It might seem that in order to do this we have to provide him with a Cartesian ego and a way of perceiving it. But it takes little reflection to see that that is wrong. Providing his language of thought with *de se* representations is just a matter of including a primitive singular term in his language of thought and then wiring him up in such a way that sensory input results in his having *de se* beliefs, and rational deliberation results in *de se* intentions. There is nothing mysterious about this. It is all a matter of programming (or wiring). In particular, we do not have to equip Oscar with "ghost" in his machine.

Belief formation

The purpose of having a language of thought is to mediate Oscar's behavioral response to his environment. Sensory input results in behavioral output, and an important part of the connection is provided by thought. The thought processes constitute reasoning and are governed by rules for reasoning – both pure reasoning and practical reasoning. The rules for pure reasoning constitute epistemic norms. In effect, epistemic norms comprise a “program” for the manipulation of sentences in the language of thought in response to sensory input.

Insofar as a cognitive agent's reasoning conforms to his epistemic norms, his beliefs are justified. At this point, it might be wondered why it is possible for human beings to reason unjustifiedly. Why not just “hardwire” a cognitive agent in such a way that he *must* reason in conformance with his epistemic norms? The answer to this seems to have to do with the existence of multiple systems of information processing. *Intellection* is the process whereby we indulge in explicit reasoning and form conclusions on that basis. But intellection is slow and consumes large amounts of our limited computational capacity. To get around this we also have a number of “quick-and-dirty” systems that allow us to form conclusions or to respond to environmental input quickly in cases in which we do not have time to deliberate. These systems are quick but “dirty” in the sense that they sometimes make mistakes in ways that can only be corrected through explicit reasoning. To illustrate, suppose you are dealt four cards that are colored on one side and numbered on the other. The exposed faces of the cards contain a two, a three, a red face, and a black face. To test the hypothesis that all cards with even numbers on one side are black on the other, which cards must you turn over? Almost all subjects say initially that you must turn over the card with a two showing and the card with a black face showing. But a little logical reflection indicates that it is not the card with the black face that you must turn over, but rather the card with the red face. Intellection leads to the right answer, but we have a strong initial tendency to give the wrong answer. Because the tendency to give the wrong answer is so uniform across subjects (even among logicians if they do not explicitly reason it out), this appears to illustrate the operation of a quick-and-dirty system that goes wrong in this case.³¹

Epistemic norms govern intellection, but in many cases we want to be able to form beliefs on the basis of our quick-and-dirty systems instead. If that is to be possible, it must be possible for us to circumvent our epistemic norms. That is why conformance to epistemic norms is not hardwired, and it is why we are able to hold unjustified beliefs. I think it is noteworthy, however, just how difficult it is to generate intuitive examples of people holding unjustified beliefs. That almost never happens. Except in cases driven by the quick-and-dirty systems, people almost always conform to their epistemic norms. It generally requires some kind of extreme psychological pressure to contravene those norms.

The operation of quick-and-dirty systems goes on “under the surface” – it is not introspectible. Only the output of the quick-and-dirty systems is introspectible. This contrasts with intellection, whose operation is introspectible. The reason for this difference is that intellection involves defeasible reasoning, and as we have seen, defeasible reasoning requires us to be aware of our own reasoning. By contrast, the operation of the quick-and-dirty systems does not involve the same kind of negative feedback and so does not require introspective awareness. On the other hand, intellection must always be available for the correction of the output of quick-and-dirty systems, so that output must be introspectible.

Reason, truth, and the individuation of concepts

We are now in a position to further understand the account of the individuation of concepts that was proposed above and was used to fend off epistemological relativism. According to the logical theory of concepts, all but the most basic concepts have nontrivial definitions in terms of more basic concepts, where these definitions state necessary and sufficient conditions for objects to exemplify the concepts. The simplest objection to the logical theory of concepts is that such definitions almost never exist. It is a very rare concept that can be given an informative definition of this sort. I have urged instead that concepts are individuated by their conceptual roles, where these are their roles in reasoning. This is the epistemological theory of concepts. Thus far I have argued *that* the epistemological theory of concepts is correct, but now we are in a position to see *why* it is correct.

Thought consists of the manipulation of sentences in the language of thought. What do we have to do to make it possible for Oscar to think? Correct reasoning consists of manipulating sentences in the language of thought in conformance with our epistemic norms. Thus all we have to do is to provide Oscar with epistemic norms and appropriate dispositions to conform to them. Providing Oscar with epistemic norms amounts to supplying conceptual roles for the primitive terms in his language of thought. Once we have done that, there is nothing further we need to do by way of interpreting his language of thought. In particular, there is no need to somehow “supply truth conditions.” That does not even make sense. There is nothing we could do that would constitute supplying truth conditions. There is no role for truth conditions to play in cognition. Once we have provided epistemic norms, Oscar can reason and his reasoning can mediate his behavioral response to sensory input. That is the only role there is for his language of thought to play.

It follows that the “semantics” of Oscar’s language of thought is completely described by describing the conceptual roles of the primitive terms of his language of thought. Those primitive terms express concepts, and I want to conclude from this that the same thing is true of concepts – concepts are characterized by their conceptual roles. Here, I understand the conceptual role of a concept to be the conceptual role of a term in the language of thought that expresses that concept. One must not jump to this conclusion too hastily, however. Concepts are Platonic items whereas terms in the language of thought are mental items. Concepts are constituents of propositions, and propositions are what we think when we have thoughts. Thought consists of the manipulation of sentences in the language of thought. General terms in the language of thought express concepts. This suggests that in constructing a semantics for the language of thought we begin with concepts and then attach them to terms from the language of thought. But I think that this is misleading. What is basic is the language of thought itself, and talk of concepts and propositions is to be explained in terms of it rather than the other way around. Concepts and propositions are forced into our ontology by the nature of defeasible reasoning. Defeasible reasoning requires us to be able to think about our thoughts. I must, for example, be able to recognize that one thought

was my reason for another thought. Thinking about my thoughts in this way, I must also be able to assign truth values to them so that I can judge, for instance, that my reason for one thought involved another thought that I now know to be false. It is possible to think about our thoughts in ways that are related only contingently to their contents. For example, if I woke up thinking that it was a beautiful day, I can think about that thought as the first thought I had this morning. But such contingent ways of thinking of thoughts are not sufficient for defeasible reasoning. In defeasible reasoning I must be able to judge that some thought I had was false, and in order to do that I must know what it was that I was thinking. For example, I cannot judge (except indirectly) that the first thought I had this morning was false unless I know that it was the thought that it was a beautiful day. In other words, I must be able to think about my thoughts in terms of their contents. This is to think about them in terms of the propositions that are thought. Thus a cognitive agent endowed with the power of defeasible reasoning must have a language of thought that contains what we can regard as ways of thinking about propositions.

The preceding suggests that defeasible reasoning requires ontologically suspicious presuppositions. But in an important sense, that is wrong. The language of thought is the vehicle for reasoning. Once we have enabled Oscar to reason by providing him with epistemic norms we have done all we need to do by way of integrating his language of thought into his cognitive machinery. In doing this, we will, among other things, have taught Oscar how to use mental representations of propositions in reasoning. But note that in enabling Oscar to reason with these mental representations we do not have to tell him what they represent – all we have to do is tell him what moves are legitimate in using them in reasoning. This is a matter of wiring or programming – not ontology. On the other hand, once we have wired him or programmed him to reason in the right ways, he then has the power to do what we call “thinking about propositions.” But in order for him to do this, we do not have to first “supply propositions for him to think about.” Looking at all this from the outside, whether there really are any propositions for him to be thinking about is totally irrelevant to his reasoning.

But next note that we can also look at this from the inside. After all, we too are cognitive agents

endowed with defeasible reasoning. We regard Oscar as thinking about propositions because he is doing what we do when we have thoughts that we regard as being about propositions. Is there something illegitimate about what we are doing? Surely not. We are conforming to our epistemic norms, and those norms make it the case that we are justified in concluding that there are propositions. This follows from the kinds of judgments we must be able to make in defeasible reasoning. For instance, I must be able to conclude that I assumed something false in reasoning in a certain way. The “something false” was a proposition. That is just what we mean by “proposition.” It follows (in accordance with my epistemic norms) from the fact that I assumed something false that *there was* something false, i.e., there was a false proposition, and hence there was a proposition. Thus belief in propositions is mandated by our epistemic norms, which are in turn constitutive of our conceptual framework. Hence we cannot be faulted for believing in propositions and concepts. We are completely justified in such beliefs.

My conclusion is that an ontology of propositions and concepts is forced upon us by our epistemic norms, or equivalently, by the semantics of our language of thought. That semantics is constituted by our epistemic norms. Concepts are, in effect, created to be the contents of general terms in the language of thought. The general terms are basic, and concepts are to be understood in terms of them rather than the other way around. Those general terms are characterized by their conceptual roles, so the concepts are equally characterized by their conceptual roles. This I take it is the ultimate and deep explanation for the epistemological theory of concepts.

If we do not need truth conditions for the functioning of concepts, it might be wondered why we even have the concept of truth. The answer is that this concept is required for defeasible reasoning. We can *think* a thought, and we can also *think about* a thought. To think a thought is to think about whatever the thought is about. On the other hand, to think about the thought is to think about a proposition. In what we might call “first-order reasoning” we just think thoughts, reasoning from one to the next without thinking about the thoughts. But in correcting defeasible reasoning we must also think about our thoughts, judging, for instance, that something we used in getting to a particular conclusion was false. We do not need the concept of truth in order to affirm a thought

while thinking that thought. The affirmation is part of the thinking. But in order to affirm a thought while thinking about it, we do need the concept of truth. The ability to ascend a level and think about our thoughts is required for the operation of defeasible reasoning, and that in turn requires that we have the evaluative concepts of truth and falsity.

The way in which we ascend levels and ascribe truth values to propositions is dictated by our epistemic norms. If *P* is a proposition I am able to entertain, then my language of thought must contain a mental representation “*P*” of *P*, and my epistemic norms must license reasoning something like the following:

What I believed was “*P*”

P

Therefore, what I believed was true.³²

Just to have a label, I will call this *disquotational reasoning*. I write the mental representation “*P*” using quotation marks because this is suggestive of the way the representation works. The quotation marks cannot be taken altogether seriously, because we are not forming the quotation name of a sentence but rather a mental representation of a proposition. But there is a strong and important parallel between the mental representation “*P*” and quotation names of sentences. We can refer to a sentence in terms of some property it has contingently, e.g., as “the first sentence on page 137.” In contrast to this, quotation names are noncontingent ways of referring to sentences. The rules of English dictate that enclosing a sentence in quotation marks generates a term designating that sentence and no other sentence. Similarly, we can think of a proposition in different ways. For example, I might think of a proposition under the contingent description “The first proposition entertained by Bertrand Russell on the morning of April 3, 1921.” But I will not ordinarily be able to ascribe truth or falsity to the proposition so conceived unless I know what proposition it is. To know what proposition it is is to be able to think of it in another way, “in terms of its content,” and know that the two propositions are the same. This is to think of the proposition in a “direct” fashion that is necessarily a way of thinking of that particular proposition. This point can be put more clearly in terms of the above schema of disquotational reasoning. That schema is supposed to be dictated by our epistemic norms. But

in order for our epistemic norms to dictate any such schema, it must be predetermined that “*P*” designates *P*. Thus “*P*” cannot designate *P* just accidentally. “*P*” must designate *P* necessarily. (This is precisely analogous to the observation that practical reasoning requires me to have a special way of thinking of myself that is necessarily a way of thinking of myself.) “*P*” designates *P* necessarily because our epistemic norms predetermine that it does, and the way they predetermine that is by licensing the above schema of disquotational reasoning.

Thus far I have been talking about propositions, but similar observations can be made about concepts. I might think of a concept as “Immanuel Kant’s favorite concept” but that will not help me in judging whether Holly exemplifies that concept unless I know what concept it is, and knowing what concept it is involves being able to think of it in another, noncontingent, way. This corresponds to the fact that we employ something like disquotational reasoning in connection with concepts too. For instance, we may move from the observation that the ball is red to the conclusion that the ball exemplifies the concept of being red. Such reasoning requires us to be able to think about the concept in a direct noncontingent way.³³ In logical contexts, when we think of concepts and propositions we usually think of them in this direct fashion.³⁴

I have described one way in which defeasible reasoning requires the concepts of truth and falsity. Defeasible reasoning also requires them in another way. This has to do with a general kind of defeater applicable to all *prima facie* reasons. All *prima facie* reasons are subject to defeat by “reliability defeaters.” If *P* is a *prima facie* reason for *Q*, and I believe *Q* on this basis, my reasoning is defeated by the discovery that I am in some kind of circumstances *C* under which *P*’s being true is not a reliable indicator of *Q*’s being true. For example, suppose I believe that the sheet of paper before me is red on the basis of its looking red. This reasoning is defeated by the discovery that the paper is illuminated by red lights and under such circumstances something’s looking red is not a reliable indicator of its being red. The concept of reliability presupposes truth. The reliability of *P* as an indicator of *Q* under circumstances of type *C* is just the probability of *Q*’s being true under circumstances of type *C* given that *P* is true. Thus this is another way in which defeasible reasoning

requires us to be able to think about propositions and their truth.

The concept of truth is required for defeasible reasoning, but it is just one more concept in our ratiocinative arsenal. The concept of truth is characterized by its role in reasoning just like any other concept. Rather than truth being fundamental and rules for reasoning being derived from it, the rules for reasoning come first and truth is characterized by the rules for reasoning about truth.

A Naturalistic Internalism

To my mind the most serious objection (other than falsehood) to all existing epistemological theories is that they are radically incomplete. Although they might give correct descriptions of some of our epistemic norms, they provide no systematic account of epistemic justification. They do not tell us what epistemic justification is all about and they do not explain why we have the epistemic norms we have. This objection can now be met. Epistemic justification consists of holding beliefs in conformance to correct epistemic norms. But as we have seen, our epistemic norms are constitutive of the concepts we have and hence it is a necessary truth that our actual epistemic norms are correct. Thus we can give an entirely adequate analysis of epistemic justification as follows:

A person’s belief is justified if and only if he holds it in conformance to his epistemic norms.

In understanding this analysis we must distinguish between doing something in accordance with norms and doing it in conformance to the norms. The analysis proceeds in terms of the latter. To say that you act in accordance with a norm is just to say that your behavior does not violate the norm. This is compatible with your doing it for some reason unrelated to the norm. To say that you act in conformance with the norm is to say not only that you act in accordance with the norm but also that your behavior is guided by the norm. Justification requires conformance – not just accordance.

This is a naturalistic analysis of epistemic justification. Reasoning is a natural process. It is something we know how to do. To say that we know how to do it is to say that it is governed by norms. Our epistemic norms are, by definition, the norms that actually govern our reasoning. This, I claim, is

a naturalistic definition of “epistemic norm.” Of course, I have not proposed an informative logical analysis of the governance process which forms the basis of these definitions, but that should not be expected. This is a natural process that we can observe in operation, not just in reasoning but in all cases of internalized procedural knowledge, and its nature can be clarified by psychological investigations. I take it that the preceding remarks about the role of epistemic norms in a cognitive machine go some distance towards clarifying all of this. But it must be emphasized that the only clarification that can be expected here is empirical clarification. We can no more provide an informative logical analysis of the governance process than we can provide an informative logical analysis of electrons or magnetism. These are natural kinds and natural processes that we discover in the world, and their nature is revealed by empirical investigation – not logical analysis.

No doubt some philosophers will be disturbed by the fact that my analysis of epistemic justification does not characterize justified beliefs in terms of a single general property (like reliability) intrinsic to the beliefs, but instead characterizes justified beliefs in terms of the reasoning underlying them. That, however, is just the way things are. What makes a belief justified is its being supported by reasoning of an approved sort, and there is no reason to think there are general intrinsic properties of beliefs that determine whether that is possible. This is connected with the charge that internalist theories give piecemeal characterizations of epistemic justification. That is only a difficulty if there is something more to be given and hence something is being left out. To clarify this point, let us distinguish between a characterization of epistemic justification in the sense of an analysis of epistemic justification, and a characterization in the sense of an epistemological theory. I gave an analysis above. I will understand an epistemological theory, on the other hand, to be a theory that attempts to describe our epistemic norms. There is nothing piecemeal about my analysis of epistemic justification, but an epistemological theory will automatically be piecemeal. This is a consequence of the nature of reason-guiding (or more generally, action-guiding) norms. Such norms tell us that under certain circumstances we are permitted to do various things and not permitted to do other things. These norms have to be rather specific because, as we saw above, they must take as input only features of the present circum-

stances that are directly accessible to our automatic processing systems. This precludes the possibility of the norms appealing to sweeping general features of the circumstances (features such as the belief being produced by a reliable process). Compare the norms for bicycle riding. These are going to be very specific, including such things as, “If you feel yourself losing momentum then push harder on the pedal” and “If you think you are falling to the right then turn the handlebars to the right.” Epistemic norms will be equally specific, telling us things (approximately) such as “If something looks red to you and you have no reason for thinking it is not red then you are permitted to believe it is red.” There is no more reason to think that we can combine all epistemic norms into one simple general formula than there is for thinking there is a single simple formula governing the use of the pedals, the handlebars, the brakes, and so on, in bicycle riding. Action-guiding norms cannot work that way.

It is illuminating to contrast this account of epistemic norms with more conventional internalist formulas. Internalists have been inclined to say instead that our epistemic norms describe the way we *actually reason*. This claim has played an important role in internalist epistemology, because it tells us how to find out what proper epistemic norms are – just examine the way we actually reason.³⁵ But this is at least misleading. We do not always reason correctly, and what epistemic norms describe is *correct* reasoning. We might similarly be inclined to say that our bike-riding norms describe the way we actually ride a bicycle, but even when we know how to ride a bicycle we sometimes make mistakes and fail to conform to our norms – I might be distracted by a pretty girl and lose my balance. Thus we might more accurately say that our bike-riding norms describe the way we actually ride a bicycle when we do it correctly. This formulation, however, sounds vacuous. After all, riding a bicycle correctly or reasoning correctly just is to conform to the norms. This creates a real puzzle for traditional accounts of action-guiding norms. The puzzle is resolved by seeing how norms for doing something are connected with knowing how to do it. The best way to describe the connection between norms and actual behavior is to say, as I did above, that our bike-riding norms and our epistemic norms are the norms that *actually guide us* in riding bicycles and reasoning. This is similar, in a very important respect, to the more customary claim that our

epistemic norms describe the way we actually reason. In each case, norms are to be elicited from what we actually do and not from some mysterious criterion, separate from our actual behavior, that tells us what we should do. But there is also an important difference between the present formulation and the traditional formulation. The present formulation does not take our reasoning behavior at face value. It recognizes that we can reason incorrectly. That need not confound us in formulating epistemic norms because, by virtue of knowing how to reason, we know how to evaluate reasoning, and so we can recognize correct and incorrect reasoning when we see it (although not necessarily with perfect reliability). This recognition process is part of the internal “non-intellectual” process whereby our norms govern our behavior. The process is non-intellectual in the sense that it does not involve our making any conscious explicit comparison of our behavior with some explicitly formulated paradigm. The process goes on under the surface. But even though we cannot consciously monitor the process, we can make use of the results by noting that under certain circumstances we judge some behavior to be permissible and other behavior to be impermissible. On the basis of these individual (normative) observations we can try to construct a general theory of right reasoning or correct bicycle riding.

This general account of epistemic norms and epistemological theories has important implications for philosophical methodology. Epistemological theories are supposed to give general accounts of “right reasoning” – that is, they purport to describe our epistemic norms. It is a contingent psychological fact that we have the norms we have. Equivalently, it is a contingent psychological fact that we employ the conceptual framework we actually employ. Does this mean that epistemological theories are contingent? This is a rather complicated question. The answer is, “Partly ‘yes,’ and partly ‘no.’” Part of what we do in epistemology is to elicit our actual epistemic norms, and that really is a contingent matter. But our ultimate conclusions are to the effect that particular concepts have conceptual roles of certain sorts. The conceptual role of a concept is a necessary feature of that concept, so it seems that our ultimate conclusions are, if true, necessarily true. Let us take this a bit more slowly, looking at each step of what transpires in an epistemological analysis.

We begin with a question such as, “How are we justified in forming beliefs about the colors of objects?”, that is, “What are the conceptual roles of color concepts?” We begin our investigation by trying to determine how we actually make such judgments. This is a matter of eliciting the epistemic norms we actually employ. That is a question about human psychology. But this does not mean that the best way to go about answering it is by performing laboratory experiments. To illustrate, consider a simpler case. Typing is an excellent example of something we learn to do automatically. When we learn to type we internalize norms telling us what to do and then we follow those norms automatically. Now suppose we want to describe those norms. Consider the question, “What finger do you use to type a ‘w’?” We *could* try to answer that question by designing a laboratory experiment in which we observe people typing “w”s under a wide variety of circumstances, but that would be silly. There is a much easier way to do it. We can *imagine* typing a “w” and observe what we do. Touch typists find themselves using their left ring finger. How can this work as a way of eliciting our norms? After all, we are not just asking what finger a person uses on a particular occasion, and people do not always type correctly. What we want to know is what finger our typing norms prescribe using to type a “w.” The reason we can answer this question by performing our thought experiment is that there is an introspectible difference between conforming to one’s internalized norms and not conforming. It is this fact that led us to the discovery of epistemic norms in the first place. We could perversely type the “w” with our right index finger, but if we did we would know that we were not doing it the way we learned to do it. The explanation for the introspectibility of this difference is something we have already observed. Namely, it is required in order for action-guiding norms to be able to correct ongoing behavior. For example, in swinging a tennis racket I monitor my swing, I can tell at each instant whether I am too high or too low, and I can correct my swing accordingly. Thus it is important to the operation of action-guiding norms that conformance to them be introspectible.

Now consider the epistemological question, “How do we judge that something is red?”, where this is intended to be a question about our epistemic norms. Sometimes we reflect upon actual judgments we observe ourselves making. More often we *imagine* making such judgments

under normal circumstances and see what goes on. For example, suppose we are considering the hypothesis that something's looking red to us gives us a *prima facie* reason for thinking it is red. We imagine being in situations in which things look red to us and note that if there are no "intervening" considerations we will come to believe that the object is red. This is not just an observation about what actually happens. It is an observation about what we know *to do* in judging colors, that is, an observation about how our automatic processing system actually guides us in reasoning about colors. It is the introspectibility of conforming to a norm that makes this observation possible. Next, suppose it is asked whether, in acquiring justification from this *prima facie* reason, it suffices to merely believe no defeaters, or if we must instead have the positive belief that there are no true defeaters. We might imagine being in situations in which we believe no defeaters but have not given the matter any thought and so have no beliefs one way or the other about whether there are any true defeaters. Again, by introspecting on whether we would be conforming to norms by making various judgments in these imagined circumstances, we find that under such circumstances we would be conforming to our norms by judging the object to be red on the basis of its looking red to us, and so we conclude that our epistemic norms permit us to make the inference without having the belief that there are no true defeaters.

This illustrates what goes on in epistemological analysis. Our basic data concern what inferences we would or would not be permitted to make under various circumstances, real or imaginary. These data concern individual cases and our task as epistemologists is to construct a general theory that accommodates it. Epistemologists have often supposed that our epistemic rules should be, in some sense, self-evident.³⁶ I have been arguing that many of the individual bits of data on which our epistemological theory is founded will, in a certain sense, be self-evident (more accurately, introspectible). By virtue of knowing how to reason we know how to tell right reasoning when we see it, and that provides us with our data. But that does not guarantee that it will be easy to construct theories describing our epistemic norms or that such theories will be obviously right once we have them. One complication both in the use of thought experiments and in interpreting our data is that because our automatic processing system

operates in a non-intellectual way without any conscious monitoring, it need not be obvious to us what makes a particular belief justified even when it is evident to us that it is justified. Our data consist in the fact that various beliefs *are* justified – not *why* they are justified. This can be illustrated by reflecting upon the fact that we have a much better account of perceptual knowledge than we do of many other kinds of knowledge. I have urged that our being appeared to in various ways provides us with *prima facie* justification for holding beliefs about our physical surroundings. The defense of this claim assumes that our beliefs in normal perception arise psychologically from our being appeared to in various ways. This is a contingent psychological thesis and cannot be regarded as a self-evident philosophical datum. Nevertheless, we regard it as a well established psychological fact, and so have no misgivings about assuming it in constructing an account of our epistemic norms.

Contrast epistemological theories of perceptual knowledge with those of a priori knowledge. We have no very good theories of a priori knowledge despite the fact that we have no difficulty telling which beliefs are justified and which are not when we are actually doing mathematics or logic. In other words, we know how to proceed in a priori reasoning, and hence we have the same kind of basic data as in the case of perception – we can recognize some beliefs as justified and others as not. What we lack in the case of a priori knowledge is a psychological account of what is going on when we have justified beliefs. We do not know the psychological source of such beliefs, and this hampers us in the attempt to construct theories of justification. This illustrates both the way in which our basic epistemological data are self-evident and the importance of contingent nonself-evident psychological facts in the construction of epistemological theories. In an important sense, describing our actual epistemic norms is part of psychology. This does not mean that it is best carried out in the laboratory, but neither can it be denied that the results of laboratory investigations can be relevant.

It is interesting to compare the epistemological methodology I have just described with the standard methodology in linguistics. In constructing grammatical theories for their own languages, linguists typically rely upon their own intuitions regarding grammaticality. They make what they call "the competence/performance distinction," according to which it is assumed that speakers of a

language know how to judge grammaticality even if they do not always produce grammatical sentences, and then they rely upon their own ability to judge grammaticality to provide them with data concerning the grammaticality of particular sentences. They then seek general theories of grammaticality to accommodate these data. What is happening here is precisely parallel to what goes on in the construction of an epistemological theory. In each case, we have procedural knowledge. In the one case we know how to reason, and in the other case we know how to talk. Because the conformance to internalized norms is introspectible, this procedural knowledge enables us to recognize particular instances of correct reasoning and correct talking, and then we can use those particular instances as data upon which to build a general theory.

The contingent enterprise of describing our actual epistemic norms is not all there is to epistemology. From a description of our epistemic norms, we want to draw conclusions about the conceptual roles of various concepts, and that is a matter of conceptual analysis. But conceptual analysis is supposed to provide us with necessary truths. How is it possible to derive necessary truths from contingent psychological generalizations? In order to answer this question, note first that true statements about the necessary properties of things need not be necessarily true. To take a well-worn example, nine is the number of planets, and nine is necessarily such that it is odd, so it follows that the number of planets is necessarily such that it is odd; but the latter is only contingently true. This is because the necessity involved is *de re* rather than *de dicto*. Similarly, a statement describing the necessary properties of a concept must refer to the concept in some way, and if the mode of reference is only contingently a way of referring to that particular concept then even though the property ascribed to the concept is a necessary property of the concept, the resulting statement will be contingent. Applying this to epistemology, in describing epistemic norms we are describing necessary properties of concepts, but this does not mean that our epistemological pronouncements are themselves necessary truths. It depends upon how we are thinking of the concepts. For example, we might be thinking of the concept *red* under some description such as “what is ordinarily expressed by the word ‘red’ in English.” The meaning of an English word is a contingent matter, and so the claim that the concept *red*, so conceived, has such-and-such a conceptual role, will be a

contingent claim about necessary properties of concepts.

Although conceptual analyses need not be expressed by necessary truths, there will be necessary truths lurking in the wings. We can think of propositions and concepts in terms of contingent descriptions of them, but I pointed out above that disquotational reasoning requires that we also have noncontingent ways of thinking of propositions and concepts. This involves the use of mental representations that are predetermined by our epistemic norms to designate the propositions or concepts that they do designate. If you think about a concept in this direct fashion and you ascribe a necessary property to it, then your belief is necessarily true. A conceptual analysis describes necessary properties of concepts, so if the conceptual analysis is expressed by a proposition that is about the concept directly then that proposition is necessarily true. Thus conceptual analyses do generate necessary truths. But they are not a priori truths. The analyses describe the conceptual roles of concepts, and our knowledge of those conceptual roles is derived from the discovery of contingent psychological generalizations regarding what epistemic norms we employ in reasoning. Thus the ultimate issue of epistemology is necessary *a posteriori* conceptual analyses.

Naturalistic epistemology is usually associated with externalism, but the present internalism is thoroughly naturalistic and, to my mind, gives epistemology much firmer roots in psychology than do existing externalist theories. Epistemology and psychology become firmly wedded.

7 Direct Realism

Having proposed a general account of the nature of epistemic norms and epistemic justification, I will end this book with a sketch of what I regard as the correct set of epistemic norms. I have argued that (1) the correct epistemic norms must be nondoxastic, and (2) they cannot be externalist. Thus what we must have is a nondoxastic internalist theory. Specifically, our epistemic norms must license inferences directly from perceptual states to physical-object beliefs without mediation by beliefs about the perceptual states. A theory of this sort is what I have called “direct realism.” Direct realism can have a structure very much like a foundations theory. My own view is that the foundations theory I have sketched elsewhere got things almost

right.³⁷ Where it went wrong was in adopting the doxastic assumption and thereby assuming that perceptual input must be mediated by epistemologically basic beliefs. Epistemic norms must be able to appeal directly to our being in perceptual states and need not appeal to our having beliefs to that effect. In other words, there can be “half-doxastic” epistemic connections between beliefs and non-doxastic states that are analogous to the “fully doxastic” connections between beliefs and beliefs that we call “reasons.” I propose to call the half-doxastic connections “reasons” as well, but it must be acknowledged that this is stretching our ordinary use of the term “reason.” The motivation for this terminology is that the logical structure of such connections is completely analogous to the logical structure of ordinary *prima facie* reasons. That is, the half-doxastic connections convey justification defeasibly.

The treatment of reasons I have adopted elsewhere took reasons to be the propositions believed rather than the states of believing those propositions.³⁸ As long as we are concerned exclusively with reasoning from beliefs, that is unassailable. But reasoning from nondoxastic states must be described differently because nondoxastic states do not have propositional content in the same way belief states do. If I believe that I am appeared to redly then my belief state has as its content the proposition that I am appeared to redly. But if I am merely appeared to redly, that perceptual state does not similarly have a propositional content. Reasoning from the perceptual state must appeal to the perceptual state itself rather than to a (non-existent) content. When one makes a perceptual judgment on the basis of a perceptual state, I want to say that the perceptual state itself is one’s reason. For the sake of uniformity, I will say the same thing about belief states, taking the belief state rather than the content of the belief state to be the reason. This involves the following modifications to our earlier definitions of “reason” and “defeater”:

DEFINITION:

A state *M* of a person *S* is a *reason* for *S* to believe *Q* if and only if it is logically possible for *S* to become justified in believing *Q* by believing it on the basis of being in the state *M*.

DEFINITION:

If *M* is a reason for *S* to believe *Q*, a state *M** is a *defeater* for this reason if and only if it is

logically possible for *S* to be in the combined state consisting of being in both the state *M* and the state *M** at the same time, and this combined state is not a reason for *S* to believe *Q*.

Reasons are always reasons *for* beliefs, but the reasons themselves need not be beliefs. We can modify our definitions of “rebutting defeater” and “undercutting defeater” similarly. A mental state is a *prima facie reason* for a belief if and only if it is a reason for which there can be defeaters.

Direct realism can adopt the same basic structure of epistemic justification, as does the foundations theory I have outlined, with the exception that epistemologically basic beliefs are replaced by “epistemologically basic mental states,” the latter being mental states that constitute reasons for various kinds of judgments. Among these mental states will be perceptual states, memory states, and perhaps some others. For instance, direct realism can handle the problem of perception by adopting nondoxastic *prima facie* reasons such as the following:

x’s looking red to *S* is a *prima facie* reason for *S* to believe that *x* is red.

This means that the perceptual state itself is the reason, and not a belief about the perceptual state.

An initially puzzling feature of direct realism emerges from the observation that although we do not usually have beliefs about our perceptual states, we *sometimes* have such beliefs. When we do they are relevant to what perceptual judgments we are justified in making. For example, if *x* looks red to me, but I believe that *x* does not look red to me, then I am not justified in taking *x* to be red on the basis of the perceptual state. Other doxastic and nondoxastic combinations are also possible. For instance, I might believe that *x* looks red to me when in fact it does not. Reflection on such cases seems to indicate that they justify one in judging that *x* is red. These examples indicate that when we have them, beliefs about nondoxastic states take precedence over the nondoxastic states themselves in determining what we are justified in believing. This can be explained by recalling the corrective role of intellection in belief formation. Intellection is always available to correct the output of both quick-and-dirty systems and intellection itself. Ordinarily, we move directly from perceptual states to judgments about physical objects, without forming any beliefs about what perceptual states

we are in. What we are now noting is that in those unusual cases in which we do form such beliefs, the beliefs about the perceptual states take precedence over the states themselves in determining what conclusions are justified. This is a direct reflection of the corrective role of intellection. An efficient cognitive agent must be able to function mechanically for the most part by employing systems that are quick but dirty. Intellection is available by correcting the dirty systems when they go wrong. In order to do that, intellection must have the power to override other systems. Intellection functions in terms of beliefs, so beliefs about input states must be given precedence over the input states themselves.

Giving beliefs about perceptual states precedence over the perceptual states themselves can be handled by taking the above nondoxastic reason to supplement the doxastic reason I have described elsewhere rather than replacing it. That doxastic reason must now be reformulated as follows:

S's believing that x looks red to him is a prima facie reason for S to believe that x is red.

Then we adopt the following defeater for the nondoxastic reason:

S's believing that x does not look red to him is a defeater for x 's looking red to S as a prima facie reason for S to believe that x is red.

Memory can be handled analogously by supplementing the doxastic mnemonic prima facie reason described elsewhere³⁹ with the following nondoxastic mnemonic prima facie reason and accompanying defeater:

S's seeming to remember P is a prima facie reason for S to believe P ; S's believing that he does not seem to remember P is a defeater for this prima facie reason.

This is only a crude sketch of the structure of reasons embodied in a plausible formulation of direct realism, but it is enough to give the flavor of such a theory.⁴⁰ It strikes me as a very plausible

kind of theory. It reins the attractive intuitions about the connection between justification and reasoning that are part and parcel of classical foundations theories, while avoiding the shortcomings of such theories by giving up the doxastic assumption. Perhaps it is good to close by considering what we can say about the skeptical problem of how we can know that we are not brains in vats with all our perceptual experiences provided by a sophisticated computer. This skeptical problem can be seen to have a trivial solution. My perceptual experience provides me with good reasons for believing that I am in my study, typing on my computer and occasionally gazing out at the mountains. It is a necessary truth that my perceptual experiences provide me with such reasons. This necessary truth is a reflection of the necessary features of my concepts that make them the concepts they are. The structure of reasons is constitutive of the concepts. That I am in my study, typing and looking out at the mountains, entails and provides a conclusive reason for believing that I am not a brain in a vat. Thus I have perfectly ordinary reasons for thinking that I am in normal surroundings and hence am not a brain in a vat. The mere logical possibility of a brain in a vat does not defeat reasons. If I had some concrete reason for thinking I really was a brain in a vat, then I would have to take that possibility seriously and I could not lay the specter of skepticism to rest so easily. But in fact, I have no such reason and hence need not be seriously concerned about the skeptical hypothesis.

This resolution of the skeptical problem may seem unsatisfying to one enamored of skeptical dilemmas. But recall that our task is not that of proving the skeptic wrong. I take it that we know from the start that the skeptic is wrong. What is wanted is an explanation of *how* we can know that we are not brains in vats rather than a proof (satisfactory to the skeptic) *that* we are not. The general explanation is that perceptual knowledge is acquired on the basis of prima facie reasons. Those reasons do not have to logically entail what they are reasons for. Our being appeared to in various ways does not logically entail that we are not brains in vats, but it does justify us in believing that we are not. End of story.

Notes

- 1 See for example Isaac Levi (1967), Keith Lehrer (1974), p. 146ff., and p. 204ff., and (1981), p. 75ff., and Alvin Goldman (1981), pp. 27–52.
- 2 See for example Hilary Kornblith (1983). See also William Alston (1978), Roderick Firth (1978), John Heil (1983), and J. Meiland (1980).
- 3 Alvin Goldman (1981) and (1986) seems to be one of the few externalists who is clear on this distinction. He distinguishes between two senses of “epistemic justification” (see section five) and adopts belief externalism with regard to one and norm externalism with regard to the other.
- 4 We can also make “third-person evaluations” of our own past behavior, but that is different from what I am calling “first-person uses” of norms.
- 5 Many philosophers appear to adopt the intellectualist model, although it is doubtful that any of them would seriously defend it if challenged. For example, Alvin Goldman (1981) appears to assume such an account of epistemic norms. The intellectualist model pervades Hilary Kornblith’s (1983) discussion. Unfortunately, it is also prominent in my own (1979) discussion.
- 6 This point has been made several times. I made it in my (1974), and James Van Cleve (1979) made it again. Despite this, I do not think that epistemologists have generally appreciated its significance. (At least, I did not.)
- 7 There has been a lot of recent work in psychology concerning human irrationality. Psychologists have shown that in certain kinds of epistemic situations people have an almost overpowering tendency to reason incorrectly. (Much of the psychological material can be found in Daniel Kahneman, Paul Slovic, and Amos Tversky (1982), and R. E. Nisbett and L. Ross (1980).) It might be tempting to conclude from this that, contrary to what I am claiming, people do not know how to reason. The short way with this charge is to note that if we did not know how to reason correctly in these cases, we would be unable to discover that people reason incorrectly. To say that we know how to reason is to invoke a competence/performance distinction. It in no way precludes our making mistakes. It does not even preclude our almost always making mistakes in specific kinds of reasoning. All it requires is that we can, in principle, discover the errors of our ways and correct them. (This is pretty much the same as the assessment offered by Jonathan Cohen, (1981). See also the critique in Alvin Goldman (1986).)
- 8 See, for example, Newell (1972), (1973), (1980), Newell and Simon (1972), and John Anderson (1976), (1983).
- 9 It might be insisted that this is at least sometimes a misleading way of talking – if our norms for doing X tell us to do Y whenever we *think* it is the case that C, we might better describe our norms as telling us to do Y when it *is* the case that C. I do not care if one chooses to talk that way, but it must be realized that it has the consequence that although the reformulated norm says to do Y when it is the case that C, knowing how to do X will really only result in our doing Y when we *think* it is the case that C. This will be important. (And, of course, norms appealing to internal states other than beliefs could not be reformulated in this manner anyway.)
- 10 It would also be a wholly implausible theory. We do not invariably have beliefs about the reliability of our inferences whenever we make them, and if norms *requiring* us to have such beliefs also require those beliefs to be justified then they lead to an infinite regress.
- 11 We do not ordinarily have any beliefs at all about the probabilities of what we believe. Furthermore, even if we did they would presumably not render our beliefs justified unless the probability beliefs were themselves justified, so we would be threatened by an infinite regress.
- 12 The *by* relation is what Alvin Goldman (1976) calls *level-generation*.
- 13 Alternatively, we may have the same norms but your physical skills make you better able to conform to them.
- 14 They may serve as recommendations in an indirect fashion by conveying to a person that there are relevant facts of which he is not apprised.
- 15 This was apparently first noted by Lewis Carroll.
- 16 This view of concepts is reminiscent of the verification theories of the logical positivists. I first defended a theory of this sort in my (1968), and in more detail in my (1974), although in those publications I talked about “justification conditions” rather than conceptual roles, and used the term a bit more narrowly. This view of concepts is also related to the somewhat cruder views expressed by Michael Dummet (1975), (1976) and Hilary Putnam (1979), (1984).
- 17 The conclusion that if different people employ different epistemic norms then they employ different concepts may seem puzzling because it appears to make it inexplicable how such people could communicate with each other. But two points should be made here. First, I doubt that there really is any variation in epistemic norms from person to person. I suspect that epistemic norms are species specific. But even if that conjecture is false, it need create no difficulty for communication. I have

- argued at length in two recent books that concepts play only an indirect role in communication. (My entire theory of language is developed in my (1982). A briefer sketch of the theory can be found in chapter two of my (1984). The reader who is concerned with this question should consult those books.)
- 18 This term comes from Jerry Fodor (1975).
 - 19 The "B-box" metaphor is due to Stephen Schiffer (1981).
 - 20 Oscar I is pretty much the same as the machines discussed by Hilary Putnam (1960).
 - 21 Putnam (1960) overlooks this.
 - 22 See Rudolf Carnap (1967), Nelson Goodman (1951), and C. I. Lewis (1946).
 - 23 Philosophers of science have long been puzzled by the role of simplicity in scientific confirmation. When two theories would each explain the data but one is significantly simpler than the other, we take the simpler one to be confirmed. But this is puzzling. What has simplicity got to do with truth? I think that the explanation for the role simplicity plays in confirmation probably lies in the kinds of considerations I have been describing. Its importance has to do with minimizing computational complexity, and its legitimacy has to do with the fact that, in a sense, the objects the generalizations are about are "free floating" and can be adjusted to minimize complexity. This is a bit vague, but I find it suggestive.
 - 24 There is an interesting purely formal question here that, I guess, lies in the domain of theoretical computer science. That is the question of the extent to which and the circumstances under which computational complexity can be decreased by introducing such "intervening variables." It is obvious that this can sometimes be achieved, but it would be interesting to have a general account of it.
 - 25 Oscar would be more like human beings if we supplied him with a "preprocessor" that modifies the input from his perceptual sensors in accordance with simple generalizations he has made about perceptual error. If he has acquired no relevant generalizations then the preprocessor will pass the input from the perceptual sensors through to his cognitive machinery unchanged, but if Oscar has acquired relevant generalizations then a red input from the perceptual sensors might be changed to orange by the preprocessor, and so on. Oscar's second-order sensors might then sense the output of the preprocessor rather than the output of the perceptual sensors themselves. This would be computationally more efficient, allowing Oscar to direct the full power of his intellect at his perceptual input only when his preprocessor cannot make sense of it. This is roughly the way people work. It involves a feedback loop from the machinery used for high-level cognitive processing to a preprocessor that lies between the sense organs and the high-level machinery.
 - 26 Recent psychologists have delighted in documenting human subjects' failure to make corrections to reasoning in light of new information. The first of these studies is apparently that of L. Ross, M. R. Lepper, and M. Hubbard (1975).
 - 27 For related accounts of mental representation in perception, see: Kent Bach (1982), Romane Clark (1973), and David Woodruff Smith (1984) and (1986).
 - 28 See my (1981) and (1982), p. 60 ff. Those publications go into some detail in describing the workings of *de re* representations, but those details are largely irrelevant to the present discussion. Related discussions occur in Diana Ackerman's (1979), (1979a), and (1980). I have proposed that a number of Keith Donnellan's (1972) well-known examples are best understood as illustrating the occurrence of *de re* representations in our thought.
 - 29 See H. N. Castaneda (1966), (1967), and (1968), Roderick Chisholm (1981), David Lewis (1979), John Perry (1977) and (1979), and my (1981) and (1982), p. 13ff. This has also played a role in some recent work in artificial intelligence. See Creary and Pollard (1985), and Rapaport (1984) and (1984a).
 - 30 The existence of *de se* beliefs raises numerous philosophical questions about the analysis of what is believed. Hector-Neri Castaneda and I have both argued that in *de se* belief one believes a proposition containing a particular kind of designator – what I previously called a "personal designator," but what might more aptly be called a "*de se* designator." Roderick Chisholm, David Lewis, and John Perry, on the other hand, have all urged that *de se* belief does not take a propositional object. They claim that the object of *de se* belief is instead a property or concept and *de se* belief involves a unique form of self-attribution. Fortunately, we need not get involved in this mare's nest at the moment. I doubt there is any substantive difference between the two accounts. Without some further constraints on what it is for something to be an object of belief, we can describe *de se* belief in either way. For example, Lewis' arguments to the contrary turn upon the assumption that any cognitive agent can, in principle, entertain any proposition. My own endorsement of *de se* propositions led me to deny that and insist instead that some propositions are logically idiosyncratic. But there appears to be no way to substantiate either position without *first* resolving the question whether belief must take a propositional object.
 - 31 The details of these experiments can be found in P. C. Wason and P. M. Johnson-Laird (1972).
 - 32 We must beware in formulating this reasoning because we are flirting with the liar paradox.

- 33 In my (1984) I argued that the distinction between modal operators and modal properties turns essentially on our having such noncontingent ways of thinking of propositions and concepts.
- 34 This is why ordinary language objections to philosophers' use of terms like "directly aware" seem beside the point. Philosophers are using these terms to express concepts they are thinking about directly. They are not thinking about those concepts under contingent descriptions like "the concept ordinarily expressed by the phrase 'directly aware' in English," and accordingly the objection that the

way they are using the term "directly aware" does not conform to ordinary usage is simply irrelevant.

- 35 Chisholm (1977) endorsed this under the label "critical cognitivism," and I endorsed it (in my (1974)) and called it "descriptivism."
- 36 This is what Ernest Sosa (1981) calls "methodism."
- 37 See my (1986), ch. 2.
- 38 *Ibid.*, esp. pp. 36–9.
- 39 *Ibid.*, ch. 2.
- 40 For more details about how to formulate a version of direct realism, see my (1974). The theory adumbrated there is a sophisticated form of direct realism.

References

- Ackermann, Diana, 1979. "Proper Names, Propositional Attitudes and Non-descriptive Connotations," *Philosophical Studies* 35, pp. 55–70.
- , 1979a. "Proper Names, Essences, and Intuitive Beliefs," *Theory and Decision* 11, pp. 5–26.
- , 1980. "Thinking about an Object: Comments on Pollock," *Midwest Studies in Philosophy*, vol. 5 (Minneapolis: University of Minnesota Press).
- Alston, William, 1978. "Meta-ethics and Meta-epistemology," in A. I. Goldman and J. Kim (eds), *Values and Morals* (Dordrecht: Reidel).
- Anderson, John, 1976. *Language, Memory, and Thought* (Hillsdale, NJ: Lawrence Erlbaum Associates).
- , 1983. *The Architecture of Cognition* (Cambridge: Harvard University Press).
- Bach, Kent, 1982. "De Re Belief and Methodological Solipsism," in Andrew Woodfield (ed.), *Thought and Object: Essays on Intentionality* (Oxford: Oxford University Press).
- Carnap, Rudolf, 1967. *The Logical Structure of the World* (London: Routledge and Kegan Paul).
- Castaneda, H. N., 1966. "He: A Study in the Logic of Self-Consciousness," *Ratio* 8, pp. 130–57.
- , 1967. "Indicators and Quasi-indicators," *American Philosophical Quarterly* 4, pp. 85–100.
- , 1968. "On the Logic of Attributions of Self-Knowledge to Others," *Journal of Philosophy* 65, pp. 439–56.
- Chisholm, Roderick M., 1957. *Perceiving: A Philosophical Study* (Ithaca, NY: Cornell University Press).
- , 1977. *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall).
- , 1981. *The First Person* (Minneapolis: University of Minnesota Press).
- Clark, Romane, 1973. "Sensuous Judgments," *Nous* 7, pp. 45–56.
- Cohen, L. Jonathan, 1981. "Can Human Irrationality Be Experimentally Demonstrated?" *The Behavioral and Brain Sciences* 4, pp. 317–70.
- Creary, L. G., and C. J. Pollard, 1985. "A Computational Semantics for Natural Language," *Proceedings of the Association for Computational Linguistics*.
- Donnellan, Keith, 1972. "Proper Names and Identifying Descriptions," in Donald Davidson and Gilbert Harman (eds), *Semantics of Natural Language* (Dordrecht: Reidel).
- Dummett, Michael, 1975. "What Is a Theory of Meaning?" in Samuel Guttenplan (ed.), *Mind and Language* (Oxford: Oxford University Press).
- , 1976. "What Is a Theory of Meaning? (II)," in Gareth Evans and John McDowell (eds), *Truth and Meaning* (Oxford: Oxford University Press).
- Firth, Roderick, 1978. "Are Epistemic Concepts Reducible to Ethical Concepts?" in A. I. Goldman and J. Kim (eds), *Values and Morals* (Dordrecht: Reidel).
- Fodor, Jerry, 1975. *The Language of Thought* (Cambridge, MA: Harvard University Press).
- Goldman, Alvin I., 1976. *A Theory of Human Action* (Princeton, NJ: Princeton University Press).
- , 1981. "The Internalist Conception of Justification," *Midwest Studies in Philosophy*, vol. 5 (Minneapolis: University of Minnesota Press).
- , 1986. *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- Goodman, Nelson, 1951. *The Structure of Appearance* (Cambridge, MA: Harvard University Press).
- , 1955. *Fact, Fiction and Forecast* (Cambridge, MA: Harvard University Press).
- Heil, John, 1983. "Believing What One Ought," *Journal of Philosophy* 80, pp. 752–65.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky, 1982. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press).
- Kornblith, Hilary, 1983. "Justified Belief and Epistemically Responsible Action," *Philosophical Review* 92, pp. 33–48.
- Lehrer, Keith, 1974. *Knowledge* (Oxford: Oxford University Press).
- , 1981. "Self-profile," in R. J. Bogdan (ed.), *Profiles: Keith Lehrer* (Dordrecht: Reidel).
- Levi, Isaac, 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science* (New York: Alfred A. Knopf).

- Lewis, C. I., 1946. *An Analysis of Knowledge and Valuation* (LaSalle, IL: Open Court).
- Lewis, David, 1979. "Attitudes De Dicto and De Se," *Philosophical Review* 87, pp. 513-43.
- Meiland, J., 1980. "What Ought We to Believe? Or the Ethics of Belief Revisited." *American Philosophical Quarterly* 17, pp. 15-24.
- Newell, Allan, 1972. "A Theoretical Exploration of Mechanisms for Coding the Stimulus," in A. W. Melton and E. Martin (eds.), *Coding Processes in Human Memory* (Washington: Winston).
- , 1973. "Production Systems: Models of Control Structures," in W. G. Chase (ed.), *Visual Information Processing* (New York: Academic Press).
- , 1980. "Reasoning, Problem Solving, and Decision Processes: The Problem Space as a Fundamental Category," in R. Nickerson (ed.), *Attention and Performance VIII* (Hillsdale, NJ: Lawrence Erlbaum Associates).
- Newell, Allan, and H. A. Simon, 1972. *Human Problem Solving* (Englewood Cliffs, NJ: Prentice-Hall).
- Nisbett, Richard, and Lee Ross, 1980. *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, NJ: Prentice-Hall).
- Perry, John, 1977. "Frege on Demonstratives," *Philosophical Review* 86, pp. 474-97.
- , 1979. "The Problem of the Essential Indexical." *Nous* 13, pp. 3-22.
- Pollock, John, 1968. "What Is an Epistemological Problem?" *American Philosophical Quarterly* 5, pp. 183-90.
- , 1974. *Knowledge and Justification* (Princeton, NJ: Princeton University Press).
- , 1979. "A Plethora of Epistemological Theories," in George Pappas (ed.), *Justification and Knowledge* (Dordrecht: Reidel).
- , 1981. "Statements and Propositions," *Pacific Philosophical Quarterly* 62, pp. 3-16.
- , 1982. *Language and Thought* (Princeton, NJ: Princeton University Press).
- , 1984. *The Foundations of Philosophical Semantics* (Princeton, NJ: Princeton University Press).
- , 1986. *Contemporary Theories of Knowledge* (Lanham: Rowman and Littlefield).
- Putnam, Hilary, 1960. "Minds and Machines," in Sidney Hook (ed.), *Dimensions of Mind* (New York: New York University Press).
- , 1979. *Meaning and the Moral Sciences* (Cambridge: Cambridge University Press).
- , 1984. *Reason, Truth, and History* (Cambridge: Cambridge University Press).
- Rapaport, W. J., 1984. "Quasi-indexical Reference in Propositional Semantic Networks," *Proceedings Coling* - 84.
- , 1984a. "Belief Representation and Quasi-indicators." *Technical Science Report 215* (SUNY Buffalo Department of Computer Science).
- Ross, L., M. R. Lepper, and M. Hubbard, 1975. "Perseverance in Self-perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigms," *Journal of Personality and Social Psychology* 32, pp. 880-92.
- Schiffer, Stephen, 1981. "Truth and the Theory of Content," in H. Parret and J. Bouverese (eds), *Meaning and Understanding* (Berlin: Walter de Gruyter).
- Smith, David Woodruff, 1984. "Content and Context of Perception." *Synthese* 61, pp. 61-87.
- , 1986. "The Ins and Outs of Perception," *Philosophical Studies* 49, pp. 187-212.
- Sosa, Ernest, 1981. "The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge," this vol., ch. 14.
- Van Cleve, James, 1979. "Foundationalism, Epistemic Principles, and the Cartesian Circle," this vol., ch. 20.
- Wason, P. C., and P. M. Johnson-Laird, 1972. *Psychology of Reasoning: Structure and Content* (London: B. T. Batsford).

A Foundherentist Theory of Empirical Justification

Susan Haack

Let us remember how common the folly is, of going from one faulty extreme into the opposite.¹

Does the evidence presented establish beyond a reasonable doubt that the defendant did it? Given the evidence recently discovered by space scientists, am I justified in believing there was once bacterial life on Mars? Is scientific evidence especially authoritative, and if so, why? Should we take those advertisements claiming that the Holocaust never happened seriously, and if not, why not? . . . Questions about what makes evidence better or worse, about what makes inquiry better or worse conducted, about disinterestedness and partiality, are of real, daily – and sometimes of life-and-death – consequence.

Of late, however, cynicism about the very legitimacy of such questions has become the familiar philosophical theme of a whole chorus of voices, from enthusiasts of the latest developments in neuroscience, to radical self-styled neo-pragmatists, radical feminists and multiculturalists, and followers of (by now somewhat dated) Paris fashions.

This cynicism is unwarranted; but dealing with it requires something a bit more radical than epistemological business-as-usual. Evidence is often messy, ambiguous, misleading, inquiry is often untidy, inconclusive, biased by the inquirers' interests; but it doesn't follow, as the cynics apparently suppose, that standards of good evidence and well-conducted inquiry are local, conventional, or

mythical. And an even half-way adequate understanding of the complexities of real-life evidence and the untidiness of real-life inquiry requires a re-examination of some of those comfortably familiar dichotomies on which recent epistemology has relied – the logical versus the causal, internalism versus externalism, apriorism versus naturalism, foundationalism versus coherentism.

Though the other dichotomies will also come under scrutiny, the main theme here will be that foundationalism and coherentism – the traditionally rival theories of justified belief – do not exhaust the options, and that an intermediate theory is more plausible than either. I call it "foundherentism."

I The Case for Foundherentism

Foundationalist theories of empirical justification hold that an empirical belief is justified if and only if it is either a basic belief justified by the subject's experience,² or else a derived belief justified, directly or indirectly, by the support of basic beliefs. Coherentist theories of empirical justification hold that a belief is justified if and only if it belongs to a coherent set of beliefs. In short, foundationalism requires a distinction of basic versus derived beliefs and an essentially one-directional notion of evidential support, while coherentism holds that beliefs can be justified only by mutual support among themselves.

The merit of foundationalism is that it acknowledges that a person's experience – what he sees, hears, etc. – is relevant to how justified he is in his beliefs about the world; its drawback is that it

Originally published in Louis Pojman (ed.), *The Theory of Knowledge: Classical and Contemporary Readings*, 2nd edn (Belmont, CA: Wadsworth, 1999), pp. 283–93.

requires a privileged class of basic beliefs justified by experience alone but capable of supporting the rest of our justified beliefs, and ignores the pervasive interdependence among a person's beliefs. The merit of coherentism is that it acknowledges that pervasive interdependence, and requires no distinction of basic and derived beliefs; its drawback is that it allows no role for the subject's experience.

Foundationlists, naturally, are keenly aware of the problems with coherentism. How could one possibly be justified in believing there's a dog in the yard, they ask, if what one sees, hears, smells, etc., plays no role? And isn't the coherentist's talk of mutual support among beliefs just a euphemism for what is really a vicious circle in which what supposedly justifies the belief that *p* is the belief that *q*, and what justifies the belief that *q* the belief that *r*... and what justifies the belief that *z* is the belief that *p*?

Coherentists, naturally, are no less keenly aware of the problems with foundationalism. What sense does it make to suppose that someone could have a justified belief that there's a dog in the yard, they ask, except in the context of the rest of his beliefs about dogs, etc.? Besides, why should we suppose that there *are* any beliefs both justified by experience alone and capable of supporting the rest of our justified beliefs? After all, foundationalists can't even agree among themselves whether the basic beliefs are about observable physical objects, along the lines of "there's a dog," or are about the subject's experience, along the lines of "it now seems to me that I see what looks like a dog" or "I am appeared to brownly." And anyway, only propositions, not events, can stand in logical relations to other propositions; so how *could* a subject's experience justify those supposedly basic beliefs?

As the two styles of theory have evolved, with each party trying to overcome the difficulties the other thinks insuperable, they have come closer together.

Strong foundationalism requires that basic beliefs be fully justified by the subject's experience; pure foundationalism requires that derived beliefs be justified exclusively by the support, direct or indirect, of basic beliefs. But weak foundationalism requires only that basic beliefs be justified to some degree by experience; and impure foundationalism, though requiring all derived beliefs to get some support from basic beliefs, allows mutual support among derived beliefs to raise their degree of justification.

Uncompromisingly egalitarian forms of coherentism hold that only overall coherence matters, so that every belief in a coherent set is equally justified. But moderated, inegalitarian forms of coherentism give a subject's beliefs about his present experience a distinguished initial status, or give a special standing to beliefs which are spontaneous rather than inferential in origin.

In a way, these moderated forms of foundationalism and coherentism lean in the right direction. But the leaning destabilizes them.

Weak foundationalism concedes that basic beliefs need not be fully justified by experience alone; but then what reason remains to deny that they could get more (or less) justified by virtue of their relations to other beliefs? Impure foundationalism concedes that there can be mutual support among derived beliefs; but then what reason remains to insist that more pervasive mutual support is unacceptable? And weak, impure foundationalism allows both that basic beliefs are less than fully justified by experience, and that derived beliefs may be mutually supportive; but now the insistence that derived beliefs can give no support to basic beliefs looks arbitrary, and the distinction of basic and derived beliefs pointless.³

Moderated, inegalitarian coherentism concedes that some beliefs are distinguished by their perceptual content or "spontaneous" origin; but isn't this implicitly to concede that justification is not after all a relation exclusively among beliefs, that input from experience is essential?

Not surprisingly, these fancier forms of foundationalism and compromising kinds of coherentism, though more sophisticated than their simpler ancestors, tend to be ambiguous and unstable. On the foundationalist side, for example, under pressure of just the kinds of difficulty my analysis identifies, C. I. Lewis moves from a pure to an impure foundationalism and then, briefly, to a kind of proto-foundherentism.⁴ And on the coherentist side, under pressure of just the kind of difficulty my analysis identifies, BonJour tries to guarantee experiential input by adding an "Observation Requirement" – which, however, is ambiguous; on one interpretation it is genuinely coherentist, but doesn't allow the relevance of experience, and on the other it allows the relevance of experience, but isn't genuinely coherentist.⁵ (BonJour now acknowledges that, after all, coherentism won't do.⁶)

Neither of the traditionally rival theories can be made satisfactory without sacrificing its distinctive

character. The obvious conclusion – though those still wedded to the old dichotomy will doubtless continue to resist it – is that we need a new approach which allows the relevance of experience to empirical justification, but without postulating any privileged class of basic beliefs or requiring that relations of support be essentially one-directional: in other words, a foundherentist theory.

II Explication of Foundherentism

The details get complicated, but the main ideas are simple.

A foundherentist account will acknowledge (like foundationalism) that how justified a person is in an empirical belief must depend in part on his experience – my version will give a role both to sensory experience, and to introspective awareness of one's own mental states. As coherentists point out, though experience can stand in causal relations to beliefs, it can't stand in logical relations to propositions. But what this shows is not that experience is irrelevant to empirical justification, but that justification is a double-aspect concept, partly causal as well as partly logical in character.

A foundherentist account will acknowledge (like coherentism) that there is pervasive mutual support among a person's justified beliefs. As foundationalists point out, a belief can't be justified by a vicious circle of reasons. But what this shows is not that mutual support is illegitimate, but that we need a better understanding of the difference between legitimate mutual support and vicious circularity – my version will rely on an analogy between the structure of evidence and a crossword puzzle.

Of course, the viability of the foundherentist approach doesn't depend on my being completely successful in articulating it. No doubt there could be other versions of foundherentism falling within these general contours but differing in their details.

I take as my starting point the following vague, but very plausible, formulation: "A is more/less justified, at *t*, in believing that *p*, depending on how good his evidence is."

By starting from here I take for granted, first, that justification comes in degrees: a person may be more or less justified in believing something. (I also assume that a person may be more justified in believing some things than he is in believing others.)

I also take for granted, second, that the concepts of evidence and justification are internally connected: how justified a person is in believing something depends on the quality of his evidence with respect to that belief.

I assume, third, that justification is personal: one person may be more justified in believing something than another is in believing the same thing – because one person's evidence may be better than another's. (But though justification is personal, it is not subjective. How justified A is in believing that *p* depends on how good *his*, A's, evidence is. But how justified A is in believing that *p* doesn't depend on how good A *thinks* his evidence is; and anyone who believed the same thing on the same evidence would be justified to the same degree.)

And I assume, fourth, that justification is relative to a time: a person may be more justified in believing something at one time than at another – because his evidence at one time may be better than his evidence at another.

"A is more/less justified, at *t*, in believing that *p*, depending on how good his evidence is." The main tasks, obviously, are to explain "his evidence" and "how good." The double-aspect character of the concept of justification is already in play; for "his," in "his evidence," is a causal notion, while "how good" is logical, or quasi-logical, in character.

The concept of justification is causal as well as logical across the board⁷ – its causal aspect is not restricted to experiential evidence alone. Quite generally, how justified someone is in believing something depends not only on *what* he believes, but on *why* he believes it. For example: if two people both believe the accused is innocent, one because he has evidence that she was a hundred miles from the scene of the crime at the relevant time, the other because he thinks she has an honest face, the former is more justified than the latter. In short, degree of justification depends on the quality of the evidence that actually causes the belief in question.

The word "belief" is ambiguous: sometimes it refers to a mental state, someone's believing something [an S-belief],⁸ sometimes it refers to the content of what is believed, a proposition [a C-belief]. "A's evidence" needs to be tied somehow to what causes A's S-belief, but must also be capable of standing in logical or quasi-logical relations to the C-belief, the proposition believed.

The idea is to begin by characterizing A's S-evidence with respect to p – this will be a set of states of A causally related to his S-belief that p ; and then to use this as the starting point of a characterization of A's C-evidence with respect to p – this will be a set of propositions capable of standing in logical or quasi-logical relations to the C-belief that p .

If A initially came to believe that the rock-rabbit is the closest surviving relative of the elephant because a fellow-tourist told him he read this somewhere, and later still believes it, but now because he has learned all the relevant biological details, he is more justified at the later time than at the earlier. So, if they are different, "A's S-evidence with respect to p " should relate to the causes of A's S-belief that p at the time in question rather than to what prompted it in the first place.

What goes on in people's heads is very complicated. There will likely be some factors inclining A towards believing that p , and others pulling against it. Perhaps, e.g., A believes that Tom Grabit stole the book because his seeing Grabit leave the library with a shifty expression and a suspicious bulge under his sweater exerts a stronger positive pull than his belief that it is possible that Tom Grabit has a light-fingered identical twin exerts in the opposite direction. Both sustaining and inhibiting factors are relevant to degree of justification, so both will be included in A's S-evidence.

In this vector of forces [the causal nexus of A's S-belief that p], besides A's present experience and present memory traces of his past experience, and other S-beliefs of his, such factors as his wishes, hopes, and fears will often play a role. But A's desire not to believe ill of his students, say, or his being under the influence of alcohol, though they may affect whether or with what degree of confidence he believes that Grabit stole the book, aren't themselves part of his evidence with respect to that proposition.

So "A's S-evidence with respect to p " will refer to those experiential and belief-states of A's which belong, at the time in question, to the causal nexus of A's S-belief that p . The phrase "with respect to" signals the inclusion of both positive, sustaining, and negative, inhibiting, evidence [respectively, A's S-evidence for p , and A's S-evidence against p]. A's S-evidence with respect to p will include other beliefs of his [A's S-reasons with respect to p]; and his perceptions, his introspective

awareness of his own mental goings-on, and memory traces of his earlier perceptual and introspective states [A's experiential S-evidence with respect to p].

The part about memory needs amplifying. A's experiential S-evidence may include present memory traces of past experience – such as his remembering seeing his car-keys on the dresser. This corresponds to the way we talk of A's remembering seeing, hearing, reading, etc. We also talk of A's remembering that p , meaning that he earlier came to believe that p and has not forgotten it. How justified A is in such persisting beliefs will depend on how good his evidence is – his evidence at the time in question, that is. A person's evidence for persisting beliefs will normally include memory traces of past perceptual experience; my belief that my high-school English teacher's name was "Miss Wright," for instance, is now sustained by my remembering hearing and seeing the name used by myself and others.

Testimonial evidence, in a broad sense – what a person reads, what others tell him – enters the picture by way of his hearing or seeing, or remembering hearing or seeing, what someone else says or writes. Of course, A's hearing B say that p won't contribute to his, A's, believing that p , unless A understands B's language. But if A believes that p in part because B told him that p , how justified A is in believing that p will depend in part on how justified A is in thinking B honest and reliable. But I anticipate.

A's S-evidence with respect to p is a set of states of A causally related to his S-belief that p . But in the part of the theory that explains what makes evidence better or worse, "evidence" will have to mean "C-evidence," and refer to a set of propositions. The two aspects interlock: A's C-evidence with respect to p will be a set of propositions, and how good it is will depend on those propositions' logical or quasi-logical relations to p ; but *which* propositions A's C-evidence with respect to p consists of depends on which of A's S-beliefs and perceptual, etc., states belong to the causal nexus of the S-belief in question.

A's C-reasons with respect to p , obviously enough, should be the C-beliefs, i.e., the propositions, which are the contents of his S-reasons. For example, if one of A's S-reasons with respect to p is his S-belief that female cardinal birds are brown, the corresponding C-reason will be the proposition that female cardinal birds are brown.

But what about A's experiential C-evidence? My proposal is that "A's experiential C-evidence with respect to p " refers to propositions to the effect that A is in the perceptual/introspective/memory states which constitute his experiential S-evidence with respect to p . Since a perceptual, etc., state can't be part of the causal nexus of A's S-belief that p unless A is *in* that state, these propositions are all true. But they need not be propositions that A believes.⁹

So A's experiential C-evidence has a distinctive status. A's C-reasons may be true or may be false, and A may be more or less justified, or not justified at all, in believing them. But A's experiential C-evidence consists of propositions all of which are, *ex hypothesi*, true, and with respect to which the question of justification doesn't arise. (This is the foundherentist way of acknowledging that the ultimate evidence for empirical beliefs is experience – very different from the forced and unnatural way in which foundationalism tries to acknowledge it, by requiring basic *beliefs* justified by experience alone.)

In line with the way we ordinarily talk about the evidence of the senses – "Why do I think there's a cardinal in the oak tree? Well, I can see the thing; that distinctive profile is clear, though the light's not too good, and it's quite far away, so I can't really see the color" – I suggest a characterization of A's experiential C-evidence in terms of propositions to the effect that A is in the sort of perceptual state a normal subject would be in when seeing this or that in these or those circumstances. For example, if A's experiential S-evidence with respect to p is his perceptual state, its looking to him as it would to a normal observer seeing a female cardinal bird at a distance of forty feet in poor light, the corresponding experiential C-evidence will be a proposition to the effect that A is in the kind of perceptual state a normal observer would be in when looking at a female cardinal bird in those circumstances.

Built into my account of experiential evidence is a conception of perception as, in a certain sense, direct. This is not to deny that perception involves complicated neurophysiological goings-on. Nor is it to deny that the judgments causally sustained by the subject's experience are interpretative, that they depend on his background beliefs as well – which, on the contrary, is a key foundherentist thought. It is only to assert that in normal perception we interact with physical things and events around us, which look a certain way

to all normal observers under the same circumstances.

You may be wondering why I include the subject's sensory and introspective experience as evidence, but not, say, his extra-sensory perceptual experience. Well, the task here is descriptive – to articulate explicitly what is implicit when we say that A has excellent reasons for believing that p , that B is guilty of wishful thinking, that C has jumped to an unjustified conclusion, and so on. As those phrases "excellent reasons" and "guilty of wishful thinking" indicate, his other beliefs should be included as part of a subject's evidence, but his wishes should not. Actually, I think it most unlikely there is such a thing as ESP; but it is excluded because – unlike sensory experience, for which we even have the phrase, "the evidence of the senses" – it has no role in the implicit conception of evidence I am trying to make explicit.

The concepts of better and worse evidence, of more and less justified belief, are evaluative; so, after the descriptive task of explication, there will be the ratificatory question, whether our standards of better and worse evidence really are, as we hope and believe they are, indicative of truth. But that comes later.

The present task is to explicate "how good" in "how good A's C-evidence is." What factors raise, and what lower, degree of justification?

Foundationalists often think of the structure of evidence on the model of a mathematical proof – a model which, understandably, makes them leery of the idea of mutual support. My approach will be informed by the analogy of a crossword puzzle – where, undeniably, there is pervasive mutual support among entries, but, equally undeniably, no vicious circle. The clues are the analogue of experiential evidence, already-completed intersecting entries the analogue of reasons. As how reasonable a crossword entry is depends both on the clues and on other intersecting entries, the idea is, so how justified an empirical belief is depends on experiential evidence and reasons working together.

Perhaps needless to say, an analogy is only an analogy, not an argument. Its role is only to suggest ideas, which then have to stand on their own feet. And there are always disanalogies; there will be nothing in my theory analogous to the solution to today's crossword which appears in tomorrow's newspaper, for instance, nor any analogue of the designer of a crossword.

But the analogy does suggest a very plausible multi-dimensional answer to the question, what makes a belief more or less justified? How reasonable a crossword entry is depends on how well it is supported by the clue and any already-completed intersecting entries; how reasonable those other entries are, independent of the entry in question; and how much of the crossword has been completed. How justified A is in believing that p , analogously, depends on how well the belief in question is supported by his experiential evidence and reasons [supportiveness]; how justified his reasons are, independent of the belief in question [independent security]; and how much of the relevant evidence his evidence includes [comprehensiveness].

On the first dimension, A's C-evidence may be conclusive for p , conclusive against p , supportive-but-not-conclusive of p , undermining-but-not-conclusive against p , or indifferent with respect to p /with respect to not- p .

Foundationalists often take for granted that evidence is conclusive just in case it deductively implies the proposition in question; but this isn't quite right. Inconsistent premisses deductively imply any proposition whatever; but inconsistent evidence isn't conclusive evidence for anything – let alone conclusive evidence for everything! Think, for example, of a detective whose evidence is: the murder was committed by a left-handed person; either Smith or Brown did it; Smith is right-handed; Brown is right-handed. Though this deductively implies that Smith did it, it certainly isn't conclusive evidence for that belief (let alone conclusive evidence for the belief that Smith did it *and* conclusive evidence for the belief that Brown did it *and* conclusive evidence for the belief that extra-terrestrials did it!).

Deductive implication is necessary but not sufficient for conclusiveness. Evidence E is conclusive for p just in case the result of adding p to E [the p -extrapolation of E] is consistent, and the result of adding not- p to E [the not- p -extrapolation of E] is inconsistent. E is conclusive against p just in case its p -extrapolation is inconsistent and its not- p -extrapolation consistent. But if E itself is inconsistent, both its p -extrapolation and its not- p -extrapolation are also inconsistent, so E is indifferent with respect to p .

Often, though, evidence is not conclusive either way, nor yet inconsistent and hence indifferent, but supports the belief in question, or its negation, to some degree. Suppose the detective's evidence

is: the murder was committed by a left-handed person; either Smith or Brown did it; Smith is left-handed; Brown is left-handed; Smith recently saw the victim, Mrs Smith, in a romantic restaurant holding hands with Brown. Though not conclusive, this evidence is supportive to some degree of the belief that Smith did it – for, if he did, we have some explanation of why.

The example suggests that supportiveness depends on whether and how much adding p to E makes a better explanatory story. But a better explanatory story than what? Conclusiveness is a matter of the superiority of p over its negation with respect to consistency. But if p is potentially explanatory of E or some component of E, it is not to be expected that not- p will be too. So I construe supportiveness as depending on the superiority of p over its rivals with respect to explanatory integration; where a rival of p is any proposition adding which to E improves its explanatory integration to some degree, and which, given E, is incompatible with p .

The word "integration" was chosen to indicate that E may support p either because p explains E or some component of E, or vice versa – that there is "mutual reinforcement between an explanation and what it explains."¹⁰ (So the concept of explanatory integration is closer kin to the coherentist concept of explanatory coherence than to the foundationalist concept of inference to the best explanation.)

Usually, as conclusiveness of evidence is taken to be the province of deductive logic, supportiveness of evidence is taken to be the province of inductive logic. But at least if "logic" is taken in its now-usual narrow sense, as depending on form alone, this looks to be a mistake. Explanation requires generality, kinds, laws – a motive for the murder, a mechanism whereby smoking causes cancer, and so forth. If so, explanatoriness, and hence supportiveness, requires a vocabulary which classifies things into real kinds; and hence depends on content, not on form alone. (Hempel drew the moral, many years ago now, from the "grue" paradox.¹¹) But there is supportive-but-not-conclusive evidence, even if there is no formal inductive logic.

Supportiveness alone does not determine degree of justification, which also depends on independent security and comprehensiveness. Suppose our detective's evidence is: the murder was committed by a left-handed person; either Smith or Brown did it; Smith is right-handed, but Brown left-

handed. The detective's evidence is conclusive that Brown did it; nevertheless, he is not well-justified in believing this unless, among other things, he is justified in believing that the murder was committed by a left-handed person, that either Smith or Brown did it, etc.

The idea of independent security is easiest to grasp in the context of the crossword analogy. In a crossword, how reasonable an entry is depends in part on its fit with intersecting entries, and hence on how reasonable those entries are, independently of the entry in question. Similarly, how justified a person is in believing something depends in part on how well it is supported by his other beliefs, and hence on how justified he is in believing those reasons, independently of the belief in question.

It is that last phrase – in my theory as with a crossword puzzle – that averts the danger of a vicious circle. The reasonableness of the entry for 3 down may depend in part on the reasonableness of the intersecting entry for 5 across – independent of the support given to the entry for 5 across by the entry for 3 down. Similarly, how justified A is in believing that *p* may depend in part on how justified he is in believing that *q* – independent of the support given his belief that *q* by his belief that *p*.

And, though “justified” appears on the right-hand side of the independent security clause, there is no danger of an infinite regress – any more than with a crossword puzzle. As in the case of a crossword eventually we reach the clues, so with empirical justification eventually we reach experiential evidence. And experiential C-evidence does not consist of other C-beliefs of the subject, but of propositions all of which are, *ex hypothesi*, true, and with respect to which the question of justification doesn't arise. This is not to deny that, as crossword clues may be cryptic, experiential evidence may be ambiguous or misleading; on the contrary, my account of experiential C-evidence is intended to recognize that it often is. It is only to say that the question of justification arises with respect to a person's beliefs, but not with respect to his experiences.

As how reasonable a crossword entry is depends not only on how well it is supported by the clue and other intersecting entries, and on how reasonable those other entries are, but also on how much of the crossword has been completed, so degree of justification depends not only on supportiveness and independent security, but also on comprehensiveness – on how much of the relevant evidence the subject's evidence includes.

Comprehensiveness promises to be even tougher to spell out than supportiveness and independent security; the crossword analogy isn't much help here, and neither is the nearest analogue in the literature, the total evidence requirement on inductions, which refers, not to the totality of relevant evidence, but to the totality of relevant available evidence – and then there is the further problem that relevance itself comes in degrees.

I am assuming, however, that (degree of) relevance is an objective matter. Naturally, whether I think your handwriting is relevant to your trustworthiness depends on whether I believe in graphology; but whether it *is* relevant depends on whether graphology is *true*.

As this reveals, though relevance, and hence comprehensiveness, is objective, judgments of relevance, and hence judgments of comprehensiveness, are perspectival, i.e., they depend on the background beliefs of the person making them. The same goes for judgments of supportiveness and independent security. How supportive you or I judge E to be with respect to *p*, for example, will depend on what rivals of *p* we happen to be able to think of; but how supportive E *is* of *p* does not. Quality of evidence is objective, but judgments of quality of evidence are perspectival.

Because quality of evidence is multi-dimensional, we should not necessarily expect a linear ordering of degrees of justification; e.g., A's evidence with respect to *p* might be strongly supportive but weak on comprehensiveness, while his evidence with respect to *q* might be strong on comprehensiveness but only weakly supportive. Nor, *a fortiori*, does it look realistic to aspire to anything as ambitious as a numerical scale of degrees of justification. But something can be said about what is required for A to be justified to *any* degree in believing that *p*.

One necessary condition is that there *be* such a thing as A's C-evidence with respect to *p*. If A's S-belief that *p* is caused simply by a blow to the head, or by one of those belief-inducing pills philosophers are fond of imagining, A isn't justified to any degree in believing that *p*. Since it is the justification of empirical beliefs that is at issue, another necessary condition is that A's C-evidence should include some experiential C-evidence – present experiential evidence, or memory traces of what he earlier saw, heard, read, etc. This is my analogue of Bonjour's Observation Requirement, obviously much more at home in

foundherentism than his requirement was in his coherentist theory. (It is not meant to rule out the possibility that some of a person's beliefs may not be sustained directly by experiential evidence, not even by memory traces, but rely on other beliefs and their experiential evidence – as in an unconventional crossword some entries might have no clues of their own but rely on other entries and their clues.¹²) A third necessary condition is that A's C-evidence with respect to p should meet minimal conditions of supportiveness, independent security, and comprehensiveness; e.g., it should be better than indifferent in terms of supportiveness. Jointly, these necessary conditions look to be sufficient.

What about the upper end of the scale? Our ordinary use of phrases like "A is completely justified in believing that p " is vague and context-dependent, depending *inter alia* on whether it is A's particular business to know whether p , and how important it is to be right about whether p ; perhaps it also runs together strictly epistemological with ethical concerns. This vague concept [complete justification] is useful for practical purposes – and for the statement of Gettier-type paradoxes. In other philosophical contexts, however, "A is completely justified in believing that p " is used in a context-neutralized, optimizing way, requiring conclusiveness, maximal independent security, and full comprehensiveness of evidence [COMPLETE justification].

The account sketched here has been personal, i.e., focussed firmly on our friend A. But this is not to deny that in even the most ordinary of our everyday beliefs we rely extensively on testimonial evidence. And where the sciences are concerned, reliance on others' evidence – and hence on the interpretation of others' words and judgments of others' reliability – is absolutely pervasive. (This reveals that not only the social sciences but also the natural sciences presuppose the possibility of interpreting others' utterances: think, e.g., of an astronomer's reliance on others' reports of observations.)

Anyway, thinking about evidence in the sciences prompts me to ask whether it is possible to extrapolate from my account of "A is more/less justified in believing that p " to a concept of justification applicable to groups of people. It might be feasible to do this by starting with the degree of justification of a hypothetical subject whose evidence includes all the evidence of each member of the group, and then discount this by some measure

of the degree to which each member of the group is justified in believing that other members are competent and honest.

III The Ratification of Foundherentism

Thus far the task has been to articulate our standards of better and worse evidence, of more and less justified belief. But what do I mean by "our"? And what assurance can I give that a belief's being justified, by those standards, is any indication that it is true?

When I speak of "our" standards of better and worse evidence, I emphatically do not mean to suggest that these standards are local or parochial, accepted in "our," as opposed to "their," community. Rather, I see these standards – essentially, how well a belief is anchored in experience and how tightly it is woven into an explanatory mesh of beliefs – as rooted in human nature, in the cognitive capacities and limitations of all normal human beings.

It is sure to be objected that the evidential standards of different times, cultures, communities, or scientific paradigms differ radically. But I think this supposed variability is at least an exaggeration, and quite possibly altogether an illusion, the result of mistaking the perspectival character of judgments of evidential quality for radical divergence in standards of better and worse evidence.

Because judgments of the quality of evidence are perspectival, people with radically different background beliefs can be expected to differ significantly in their judgments of degree of justification. It doesn't follow that there are no shared standards of evidence. If we think of the constraints of experiential anchoring and explanatory integration rather than of specific judgments of the relevance, supportiveness, etc., of this or that evidence, I believe we will find commonality rather than divergence.

Again, the point is easier to see in the context of the crossword analogy. Suppose you and I are both doing the same crossword puzzle, and have filled in some long central entry differently. You think, given your solution to that long central entry, that the fact that 14 down ends in a "T" is evidence in its favor; I think, given my solution to that long central entry, that the fact that it ends in a "D" is evidence in its favor. Nevertheless, we are both

trying to fit the entry to its clue and to other already-completed entries. Now suppose you and I are both on an appointments committee. You think the way this candidate writes his "g"s indicates that he is not to be trusted; I think graphology is bunk and scoff at your "evidence." Because of a disagreement in background beliefs, we disagree about what evidence is relevant. Nevertheless, we are both trying to assess the supportiveness, independent security, and comprehensiveness of the evidence with respect to the proposition that the candidate is trustworthy.

But even if I am wrong about this, even if there really are radically divergent standards of evidential quality, it wouldn't follow that there are no objective indications of truth; *variability* of standards does not, in and of itself, imply *relativity* of standards.¹³ So those epistemic relativists who have inferred that, since judgments of justification vary from community to community, there can be no objectively correct standards of better and worse evidence, have committed a *non sequitur* as well as relying on a dubious premiss.

As for those who have succumbed to epistemic relativism because they have given up on the concept of truth, I have room here only to say that theirs seems to me an entirely factitious despair.¹⁴ In any case, all that will be required of the concept of truth in what follows is that a proposition or statement is true just in case things are as it says.

Supposing – as I believe, and so do you – that we humans are fallible, limited but inquiring creatures who live in a world which is largely independent of us and what we believe about it, but in which there are kinds, laws, regularities; and supposing – as I believe, and so do you – that our senses are a source, though by no means an infallible source, of information about things and events in the world around us, and introspection a source, though by no means an infallible source, of information about our own mental goings-on; then, if any indication of how things are is possible for us, how well our beliefs are anchored in our experience and knit into an explanatory mesh is such an indication. (And supposing – as I believe, and so, probably, do you – we have no other sources of information about the world and ourselves, no ESP or clairvoyance or etc., then this is the only indication we can have of how things are.)

That last paragraph was nothing like an *a priori* ratification of foundherentism; for those "supposing" clauses are empirical in character. Assumptions about human cognitive capacities

and limitations are *built into* our standards of evidential quality; so the truth-indicateness of those standards depends on the truth of those empirical assumptions. But neither was that last paragraph much like the appeals to psychology or cognitive science on which some epistemological naturalists of a more extreme stripe than mine propose to rely; for the assumptions referred to in my "supposing" clauses, though empirical, are of such generality as to be rather philosophical than scientific in character.

Those assumptions would surely be presupposed by any conceivable scientific experiment. But they are well integrated with what the sciences of cognition have to tell us about the mechanisms of perception and introspection, and of when and why they are more or less reliable, and with what the theory of evolution suggests about how we came to have the sort of information-detecting apparatus we do. As one would hope, the epistemological part of my crossword – the part where the entries are themselves about crosswords – interlocks snugly with other parts.

But what am I to say to those readers familiar with Descartes' failed attempt to prove "what I clearly and distinctly perceive is true," who are bound to suspect that I must be arguing in a circle? After pointing out that I have not offered a ratificatory argument in which some premiss turns out to be identical with the conclusion, nor an argument relying on a certain mode of inference to arrive at the conclusion that this very mode of inference is a good one – only that, to borrow Peirce's words, by now "the reader will, I trust, be too well-grounded in logic to mistake mutual support for a vicious circle of reasoning."¹⁵

And what am I to say to readers worried about the Evil Demon, who are bound to object that I have not ruled out the possibility that our senses are not a source of information about the external world at all? After pointing out that since, *ex hypothesi*, his machinations would be absolutely undetectable, if there were an Evil Demon *no* truth-indication would be possible for us – only that my claim is a conditional one: that, if any truth-indication is possible for us, the foundherentist criteria are truth-indicative. (I could discharge the antecedent, and arrive at a categorical conclusion, by adopting a definition of truth along Peircean lines, as the opinion that would survive all possible experiential evidence and the fullest logical scrutiny; but I prefer the more cautious, and more realist, strategy.)

Determined skeptics won't be persuaded; but determined skeptics never are! And the rest of you may notice that foundherentism enables us to sidestep another dichotomy which has – if you'll pardon the pun – bedeviled recent epistemology: *either* a hopeless obsession with hyperbolic skepticism, *or* a hopeless relativism or tribalism preoccupu-

pied with “our (local, parochial) epistemic practices.” Foundherentism, I believe, provides a more realistic picture of our epistemic condition – a robustly fallibilist picture which, without sacrificing objectivity, acknowledges something of how complex and confusing evidence can be.

Notes

This brief statement of foundherentism is based primarily on my *Evidence and Inquiry: Towards Reconstruction in Epistemology* (Oxford: Blackwell, 1993), especially chapters 1, 4, and 10. I have also drawn on material from earlier articles of mine, especially “Theories of Knowledge: an Analytic Framework,” *Proceedings of the Aristotelian Society* LXXXIII (1982–3), pp. 143–57 (where foundherentism was first introduced), “C. I. Lewis,” in *American Philosophy*, ed. Marcus Singer, Royal Institute of Philosophy Lecture Series, 19 (Cambridge: Cambridge University Press, 1985), pp. 215–39, and “Rebuilding the Ship While Sailing on the Water,” in R. Barrett and R. Gibson (eds), *Perspectives on Quine* (Oxford: Blackwell, 1990), pp. 111–27 (where some of the key ideas of foundherentism were developed). I have drawn as well on material from the symposium on *Evidence and Inquiry* published in *Philosophy and Phenomenological Research* LVI. 3 (1996), pp. 611–57, and from the debate with Bonjour in *Synthese* 112.1 (July 1997), pp. 7–35.

1 Thomas Reid, *Essays on the Intellectual Powers* (1785), in R. E. Beanblossom and K. Lehrer (eds), *Thomas Reid: Inquiry and Essays*, (Indianapolis, IN: Hackett, 1983), vol. VI, p. 4.

2 I restrict my attention here to experientialist forms of foundationalism, ignoring, e.g., foundationalist theories of a priori knowledge.

3 My characterization of foundationalism is quite standard; cf. for example, Alston's in J. Dancy and E. Sosa (eds), *Companion to Epistemology* (Oxford: Blackwell, 1992), p. 144, or Sosa's in “The Raft and the Pyramid,” this vol., ch. 14. But matters have been confused because, in “Can Empirical Knowledge Have a Foundation?,” this vol., ch. 21, and *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1986), p. 28, Bonjour uses “weak foundationalism” to refer a style of theory which is both weak and impure, in my sense, and in addition allows mutual support among basic beliefs and – apparently – allows “basic” beliefs to get support from “derived” beliefs. As my scare quotes indicate, once one-directionality has been so completely abandoned it is unclear that the theory really qualifies as foundationalist at all; certainly the basic/derived distinction has become purely

proforma. See also Haack, “Reply to Bonjour,” *Synthese* 112.1 (July 1997), pp. 25–35.

4 See *Evidence and Inquiry*, ch. 2, for details.

5 See *ibid.*, ch. 3, for details.

6 Laurence Bonjour, “Haack on Justification and Experience,” *Synthese* 112.1 (July 1997), pp. 13–15.

7 An idea I first began to work out in “Epistemology With a Knowing Subject,” *Review of Metaphysics* XXXIII.2 (1979), pp. 309–36.

8 Expressions introduced in square brackets are my new, technical terms, or special, technical uses of familiar terms.

9 So my theory is not straightforwardly externalist, since A's S-evidence must consist of states of A – states, furthermore, of which A can be aware; but neither is it straightforwardly internalist, since A's experiential C-evidence consists of propositions A need not believe or even conceive.

10 W. V. Quine and J. Ullian, *The Web of Belief* (New York: Random House, 1970), p. 79.

11 N. Goodman, “The New Riddle of Induction” (1953), in *Fact, Fiction and Forecast* (Indianapolis, IN: Bobbs-Merrill, 2nd edn, 1965), pp. 59–83; C. G. Hempel, “Postscript on Confirmation” (1964), in *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 47–52.

12 In case a desperate foundationalist is tempted to try seizing on this in hopes of salvaging the derived/basic distinction, let me point out that beliefs without direct experiential evidence could contribute to the support of beliefs with direct experiential evidence; and that this maneuver would identify no plausible *kind* of belief as basic/as derived – think, e.g., of a scientist whose belief that electrons are composed thus and so is sustained by what he sees in the bubble chamber.

13 See also Susan Haack, “Reflections on Relativism: From Momentous Tautology to Seductive Contradiction,” *Noûs* Supplement (1996), pp. 297–315, and in James E. Tomberlin (ed.), *Philosophical Perspectives, 10: Metaphysics* (Oxford: Blackwell, 1996), pp. 297–315; reprinted in Haack, *Manifesto of a Passionate Moderate: Unfashionable Essays* (Chicago: University of Chicago Press, 1998), pp. 149–66.

Susan Haack

- 14 I have more to say in "Confessions of an Old-Fashioned Prig," in *Manifesto of a Passionate Moderate*, pp. 7–30.
- 15 C. S. Peirce, *Collected Papers*, eds C. Hartshorne, P. Weiss, and A. Burks (Cambridge, MA: Harvard University Press, 1931–58), 6.315.



PART V

The Pyrrhonian Problematic

Introduction

The selections in this section each attempt to address the Pyrrhonian problematic, which was sketched in the introduction to Part III. James Van Cleve offers an externalist, foundationalist response. Laurence Bonjour presents a vigorous attack on foundationalist externalism. Ernest Sosa offers a bi-level epistemology which, while externalist and foundationalist at the level of animal knowledge, acquires an internalist, coherentist dimension at the level of reflective knowledge.

Van Cleve sees the "Cartesian circle," in its generalized form, as the problem of the criterion, which itself is a close relative of the Pyrrhonian problematic. He formulates the problem, as it applies to Descartes, as follows: I can know that whatever I clearly and distinctly perceive is true only if I first know that God exists and is not a deceiver; but I can know the latter only if I first know that whatever I clearly and distinctly perceive is true. These claims are incompatible. One must go. But which? The latter, Van Cleve argues. Textual evidence is adduced to show that this was Descartes's answer as well. Descartes attempted to establish that whatever is clearly and distinctly perceived is true by beginning with lower-level beliefs that in fact are instances of clear and distinct perception, and then reasoning from these original beliefs, in a fashion that preserves clarity and distinctness of perception, to arrive at the conclusion that whatever is clearly and distinctly perceived is true. According to Van Cleve, Descartes was right to think that if in fact whatever is clearly and distinctly perceived yields knowledge, then if clear and distinct premises clearly and distinctly lead to the conclusion that whatever is

clearly and distinctly perceived is true, then the latter can become clear and distinct, and therefore known.

In the second part of his piece, Van Cleve applies these reflections on the Cartesian circle to contemporary debates in epistemology. He argues that if in fact an epistemic principle such as Chisholm's Principle (C) is true, then fulfilling its non-epistemic antecedent is sufficient for fulfilling its epistemic consequent, independently of whether one knows the truth of the principle itself:

- (C) If there is a certain sensible characteristic *F* such that *S* believes that he perceives something to be *F*, then it is *evident* to *S* that he is perceiving something to have that characteristic *F*, and also that there is something that is *F*

Thus, a faculty may yield knowledge even if the subject lacks knowledge that the faculty generally yields knowledge. Furthermore, there is room for a kind of higher-level knowledge which takes as its object the epistemic principles in virtue of the truth of which one possesses the lower-level knowledge. Van Cleve discusses three possible ways such knowledge might be achieved: through naturalistic (scientific) methods, through critical cognitivist methods, or by immediate awareness. In the end, he finds the second method most promising. The critical cognitivist examines various beliefs, taken to be knowledge, and asks what must be the case if these beliefs in fact are to be knowledge. The method results in arguments in favor of the truth of epistemic principles. And if in

Introduction

fact the arguments are good ones – if their premises are pieces of knowledge and their conclusions are sufficiently justified by their premises – then one does indeed reach higher-level knowledge of epistemic principles.

Laurence Bonjour argues that foundationalism, whether in its givenist or in its externalist form, fails to solve the problem that it was designed to solve, viz. the Pyrrhonian problematic. He bases his argument on the very nature of epistemic justification as essentially connected to the cognitive goal of truth. What this connection amounts to, he claims, is that a belief is justified only if one has good reason to think it is true. Not only this: one must have good reason to think, regarding whatever feature that in fact makes the belief justified, that beliefs possessing that feature are likely to be true. The problem for foundationalists is that there is only one conceivable way foundational beliefs could count as justified: one would have to know, regarding whatever feature makes foundational beliefs foundational, that beliefs with that feature were likely to be true. Yet this is not knowable *a priori*, at least about any of the sources of empirical knowledge. Nor can it be known *a posteriori*, for it could only be established using circular reasoning, which is eschewed by foundationalists.

BonJour anticipates objections to his interpretation of the conditions required for having good reason to think a belief true. He replies to externalists by charging them with evasion, claiming that they waive the requirement of having good reasons arbitrarily, merely to avoid the problem. In favor of givenists, he acknowledges that the search for the given is a search for something that justifies foundational beliefs. Nonetheless, they seek the impossible. There cannot be a state of mind that is able to impart justification but needs no justification itself. To impart justification, a state must have assertive content, and having assertive content suffices for standing in need of justification.

Ernest Sosa opens with an examination of Moore's famous proof of an external world. Sosa

concludes that no progress can be made in staving off the skeptic about perceptual knowledge unless perceptual belief is regarded as non-inferential. Once we do this, we can look to the model of knowledge provided by Descartes. If a faculty is reliable, then one can use it to gain knowledge, even perhaps knowledge that the faculty is reliable. Sosa thus distinguishes between a lower grade and a higher grade of knowledge, i.e., animal and reflective knowledge. Animal knowledge is knowledge owing to the use of externally apt or reliable faculties, which Sosa calls virtues. Reflective knowledge is knowledge that results from combining the use of such externally apt virtues with explanation- and coherence-seeking reason regarding the nature and source of one's knowledge. Reflective knowledge provides what Sosa calls a "perspective" on one's epistemic functioning. Of course, one could have such a perspective but lack knowledge. The crystal-ball gazer may weave an elaborate web of comprehensively coherent belief about the accuracy of his crystal-gazing. What he lacks is the external aptness required for knowledge. Barry Stroud criticizes this externalist element in Sosa's account, charging that all an externalist can really assert are conditionalized knowledge claims: "if I am right that perception is reliable, and reliability yields knowledge, then I understand my knowledge." Sosa replies that we are under no such constraints. Insofar as we really are certain that perception is reliable and that reliability yields knowledge, we can claim that we know through perception. Hence we can and will declare outright that we have an understanding of our knowledge.

It is interesting to compare Sosa's bi-level view of animal knowledge and reflective knowledge not only with the bi-level epistemology of *cognitio* and *scientia* that he attributes to Descartes, but also with Susan Haack's foundherentism (in Part IV) and with John McDowell's proposed defense of internalism with due sensitivity to externalist insights (in Part VII).

Further Reading

- Alston, William, "Epistemic Circularity," in *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989), pp. 319–49.
- , *The Reliability of Sense Perception* (Ithaca, NY: Cornell University Press, 1993).

- Amico, Robert, *The Problem of the Criterion* (Lanham, MD: Rowman and Littlefield, 1993).
- Burnyeat, M. F. (ed.), *The Skeptical Tradition* (Berkeley: University of California Press, 1983).

- Chisholm, Roderick, "The Problem of the Criterion," in *Theory of Knowledge* (Englewood Cliffs: Prentice-Hall, 1966, 2nd edn 1977, 3rd edn 1989).
- Fogelin, Robert, *Pyrrhonian Reflections on Knowledge and Justification* (Princeton: Princeton University Press, 1994)
- Popkin, Richard, *Scepticism from Erasmus to Spinoza* (Berkeley: University of California Press, 1979).
- Sosa, E., "Philosophical Scepticism and Epistemic Circularity," *Aristotelian Society Supplementary Volume 68* (1994), pp. 263–90.
- Strawson, P. F., *Skepticism and Naturalism: Some Varieties* (London: Methuen, 1985).
- Stroud, Barry, *The Significance of Philosophical Scepticism* (Oxford: Oxford University Press, 1984).
- , "Scepticism, 'Externalism', and the Goal of Epistemology," in *Aristotelian Society Supplementary Volume 68* (1994), pp. 291–307.

Foundationalism, Epistemic Principles, and the Cartesian Circle

James Van Cleve

The problem of the Cartesian Circle is sometimes treated as though it were merely an exercise for scholars: Descartes fell into it, and their job is to get him out of it. But more is at stake than extricating Descartes. In its generalized form, the Cartesian Circle is none other than the Problem of the Criterion, a problem that any epistemology must face. Moreover, to solve the problem of the Circle one must answer questions about epistemic principles that are pivotal in contemporary debates between foundationalists and coherentists. There is reason to hope, therefore, that by examining Descartes's problem we can throw light on problems of our own.

This paper is divided into two parts. In Part One I examine solutions to the problem of the Circle that are possible within Descartes's own framework. In Part Two I show how what we learn in Part One may be used to resolve some contemporary disputes that hinge on the status of epistemic principles

Part One

I

The problem of the Cartesian Circle arose for Descartes because he appeared to commit himself to each of the following propositions:

- (1) I can know (be certain) that (p) whatever I perceive clearly and distinctly is true only if I

first know (am certain) that (q) God exists and is not a deceiver.

- (2) I can know (be certain) that (q) God exists and is not a deceiver only if I first know (am certain) that (p) whatever I perceive clearly and distinctly is true.

Obviously, if (1) and (2) are both true, I can never be certain of either p or q . To be certain of either I would already have to be certain of the other. Yet Descartes said he was certain of *both* p and q . How can this be possible?¹

Any adequate solution to the problem of the Cartesian Circle will plainly have to deny either (1) or (2). In the next section I consider a famous solution that denies (1).

II

The solution I have in mind is the Memory Gambit, according to which God is called upon to guarantee not the truth of clear and distinct perceptions, but the accuracy of our memories. The most able recent defender of this solution is Willis Doney, who cites a number of passages that seem to show that this solution was Descartes's own.² In these passages Descartes says that if I remember³ clearly and distinctly perceiving something that I do not *now* clearly and distinctly perceive, I can be certain of it if and only if I know that God exists and is not a deceiver. He also says that an atheist can know theorems of geometry *if* he is clearly and distinctly perceiving them at the time, but warns that doubts may arise later that only knowledge of God's veracity can remove. Doney concludes that the function of God is to guarantee the accuracy of

Originally published in *The Philosophical Review* 88 (1979), pp. 55-91; copyright Cornell University, reprinted by permission of the publisher and the author.

memory,⁴ and that the atheist's plight is that in his ignorance of God's guarantee he cannot be sure that he really did clearly and distinctly perceive what he remembers so perceiving. Limited to what he can clearly and distinctly perceive at each moment, his knowledge will be "meager and fugitive."

Although the Memory Gambit has some textual plausibility, Harry Frankfurt has convincingly argued that it is neither the solution Descartes intended, nor a very satisfying solution in its own right.⁵ I shall not repeat his case here, but I do want to point out an alternative explanation of the passages that make the Memory Gambit tempting. Consider the following sequence of propositions:

- (1) I remember clearly and distinctly perceiving *p*.
- (2) So, I *did* clearly and distinctly perceive *p*.
- (3) So, *p* is true.

Descartes says that the atheist cannot argue from (1) to (3). According to the Memory Gambit, this is because he cannot take the step from (1) to (2). But another possible explanation is that he cannot take the step from (2) to (3). And if this is what Descartes had in mind, then he must have felt that a divine guarantee for clear and distinct perception was needed after all.⁶

III

I pass now to solutions that deny (2). Interesting solutions of this type have been offered by Alan Gewirth and Fred Feldman. Gewirth sums up his basic strategy as follows:

Descartes's argument is not circular, for, while it is by the *psychological* certainty of clear and distinct perceptions that God's existence is proved, what God guarantees is the *metaphysical* certainty of such perceptions. [Emphasis mine]⁷

Psychological certainty is a subjective affair, implying only an irresistible compulsion to believe.⁸ Metaphysical certainty, on the other hand, is an objective affair, implying truth.⁹ If we let "certain" in (1) and (2) express metaphysical certainty, then Gewirth would deny (2). I do not need to have metaphysical certainty that clear and distinct perceptions are true before I prove God's existence; it suffices if clear and distinct perception gives me psychological certainty.¹⁰ But once I am

psychologically certain that God exists, I can use this fact to establish the metaphysical certainty of clear and distinct perceptions.

The big question prompted by this approach, of course, is this: how can mere psychological certainty about God possibly give rise to metaphysical certainty about clear and distinct perceptions? Gewirth's answer to this question will emerge from the following reconstruction of his account of Descartes's program.¹¹

- (1) I perceive clearly and distinctly that the premises of Descartes's theological arguments are true, and that their conclusions follow from them. I thereby arrive at clear and distinct perception (and thus psychological certainty) that God exists and is no deceiver.
- (2) A proposition *P* is metaphysically certain if and only if there is no proposition *R* that is a reason for doubting *P*.¹²
- (3) *R* is a reason for doubting *P* only if *R* is itself clearly and distinctly perceived (and thus something I am psychologically certain of).¹³
- (4) The only reason for doubting the truth of clear and distinct perceptions is the hypothesis that God is a deceiver.
- (5) If I clearly and distinctly perceive that God exists and is no deceiver, then I do *not* clearly and distinctly perceive that God is a deceiver.
- (6) I do not clearly and distinctly perceive that God is a deceiver. (1) and (5).
- (7) The hypothesis that God is a deceiver is not a reason for doubting anything. (3) and (6).
- (8) All clear and distinct perceptions are metaphysically certain. (2), (4), and (7).

In the beginning clear and distinct perceptions are only psychologically certain. But as soon as we have used them to prove that God exists and is no deceiver, they become metaphysically certain.¹⁴ This is because by proving the existence and veracity of God we eliminate the only possible reason for doubting clear and distinct perceptions. But according to (2), if there is no reason to doubt something, it is metaphysically certain. In this way our original psychological certainties get paralyzed into metaphysical certainties.

As it stands, this argument is open to two serious objections.¹⁵ In the first place, premise (3) sets the requirements that must be satisfied by a reason to doubt absurdly high. Descartes himself certainly does not require that reasons to doubt be clearly and distinctly perceived. He says that the reasons to doubt he brings forth are "very slight,"

“metaphysical,” and “themselves doubtful.”¹⁶ In the second place, premise (5) seems to presuppose that we can never clearly and distinctly perceive each of two mutually inconsistent propositions. But this is a thesis one would rather see as a conclusion than as a premise of the Cartesian enterprise.¹⁷

Fortunately, however, these objections can both be avoided by a single change in the argument. We need only replace premise (3) by the following premise:

- (3) *R* is a reason for doubting *P* only if it is not the case that its negation, not-*R*, is clearly and distinctly perceived. (I shall abbreviate the consequent by “*R* is not excluded by clear and distinct perceptions.”)

Here we no longer require that reasons to doubt be upheld by clear and distinct perception, but only that they not be condemned by it. This enlarges the class of permissible reasons to doubt, thus mitigating the objection to premise (3). Moreover, we may now dispense altogether with the objectionable premise (5), since the conclusion, (8), follows from (1), (2), (3), and (4) alone.

There remains, however, an objection that is fatal to Gewirth’s whole approach. The revised argument establishes only that clear and distinct perceptions are metaphysically certain *in the sense jointly defined by premises (2) and (3)*. That sense amounts to this: a proposition is metaphysically certain if and only if every reason for doubting it is excluded by clear and distinct perceptions. Now it was assumed initially that clear and distinct perceptions are only psychologically certain. What is added when at the end of the argument we say that God’s veracity and other things clearly and distinctly perceived are metaphysically certain? Just this: that we are psychologically certain not only of those propositions themselves, but also of the falsehood of every reason for doubting them. Thus, we have not advanced to a new *kind* of certainty at all. We have merely extended the class of psychological certainties.

Descartes played for higher stakes. The certainty he sought was certainty in a sense entailing both *maximal evidence* and *truth*. Despite what Gewirth says, metaphysical certainty in his sense entails neither.¹⁸ It remains at bottom a purely psychological notion.

In the hope of obtaining a conclusion that is epistemologically more nourishing, let us turn to Feldman.¹⁹ His reconstruction of Descartes is

identical in structure with the one we obtained by revising Gewirth, but there is an important difference: he replaces the concept of psychological certainty with the epistemic concept of *practical* certainty. Practical certainty is the sort of certainty involved in ordinary knowing of the justified-true-belief variety. Being practically certain of something, unlike being psychologically certain of it, entails having some justification for believing it.

Feldman’s argument may be set out as follows:²⁰

- (1) By means of Descartes’s theological arguments, I attain practical certainty that God exists and is no deceiver.
- (2) A proposition *P* is metaphysically certain if and only if there is no proposition *R* that casts metaphysical doubt on *P*.
- (3) *R* casts metaphysical doubt on *P* only if it is not the case that its negation, not-*R*, is practically certain.
- (4) The only proposition that casts metaphysical doubt on any clear and distinct perception is the hypothesis that God is a deceiver.
- (5) All clear and distinct perceptions are metaphysically certain. (1)–(4).

The logic of this argument is the same as Gewirth’s. Before we have proved the veracity of God, clear and distinct perceptions are only practically certain. They are not metaphysically certain, because doubt is cast on them by the hypothesis that God is a deceiver. But as soon as we become practically certain that God is *not* a deceiver, that hypothesis, by (3), is no longer eligible to cast doubt. By (4), however, it is the only candidate, so *nothing* any longer casts doubt on clear and distinct perceptions. Therefore, by (2), they are henceforth metaphysically certain.

In light of (2) and (3), Feldman’s conclusion is equivalent to the following: we are practically certain not only of clear and distinct perceptions, but also of the falsehood of every proposition that would cast doubt on them. This is an improvement over Gewirth’s conclusion, since it implies that clear and distinct perceptions have something going for them epistemically. But my basic misgiving remains. Feldman’s conclusion, like Gewirth’s, provides no guarantee that clear and distinct perceptions are *true*. For a Cartesian this is not enough.²¹

Gewirth and Feldman fall short of giving us what Descartes wanted for the same reason. They both define metaphysical certainty in terms of the absence of reasons to doubt. Then they make it

very hard for anything to qualify as a reason to doubt. As a result, they make it *very easy* for things to qualify as metaphysically certain. Their standards for certainty are thus set too low.

Descartes, on the other hand, was much more liberal about what could count as a reason to doubt. For him, a proposition could function as a reason to doubt just so long as it was (in Feldman's terms) a "metaphysical possibility."²² This makes his standards for certainty very high. In the next section I defend Descartes's high standards.

IV

According to H. A. Prichard, "We can only be uncertain of one thing because we are certain of something else."²³ And according to Wittgenstein, "The game of doubt presupposes certainty. If you tried to doubt everything, you would not get as far as doubting anything."²⁴ A similar point was made in Descartes's own day by Bourdin, who asked concerning the reasons Descartes gave for doubting, "If they are doubtful and replete with suspicion, how can they have brought any force to bear upon you?"²⁵

Descartes was unimpressed with this objection. In reply to Bourdin he said, "We may well enough be compelled to doubt by arguments that are in themselves doubtful."²⁶

Descartes was right about this. Reasons to doubt need not be certain: they need only be epistemically possible. To vindicate Descartes against Bourdin and company, I shall now construct an argument that purports to show that one thing is uncertain without presupposing that anything else is certain. There are three preliminaries. First, I define epistemic possibility as follows: if P is a proposition that S is considering at t , then P is *epistemically possible* for S at t if and only if S is not certain at t of not- P . Second, by the Demon Hypothesis I mean the following: an evil demon brings it about that whatever seems evident to me is false. Finally, let T be the proposition that $2 + 3 = 5$. (If there is any proposition the reader finds *more* evident than $2 + 3 = 5$, he may let T be that one instead.) Now here is the argument:

- (1) The following proposition is epistemically possible for me: T seems evident to me and the Demon Hypothesis is true.
- (2) If P entails Q and P is epistemically possible for me, then Q is epistemically possible for

me. (In other words, epistemic possibility is transmitted by entailment. This is the analogue of a theorem in modal logic.)

- (3) (T seems evident to me and the Demon Hypothesis is true) entails (T is false). So,
- (4) (T is false) is epistemically possible for me. So, by definition,
- (5) I am not certain that T is true.

Here we have a valid argument that leads to the conclusion that I am not certain that $2 + 3 = 5$. Yet nowhere in the argument is anything claimed to be certain. The Demon Hypothesis, the most powerful of Descartes's reasons to doubt, is said only to be epistemically possible.

"Ah, yes, but wait just a minute," I hear an objector say. "It is true that *in* your premises nothing is claimed to be certain. But in order for the argument to do its job, you who advance it must be certain *of* the premises, that is, you must be certain that they are true." The objector is mistaken. If the premises of my skeptical argument are merely *true*, whether known to be so or not, the conclusion will also be true. And in that case I will not be certain that $2 + 3 = 5$. But the premises could all be true without my being certain of anything. Therefore, there are conditions sufficient for my being uncertain about one thing that do not require my being certain about something else. And this is just what the argument was supposed to illustrate.

Nevertheless, there is another and more cogent objection to the argument. Premise (2) is false, and derives only spurious support from the analogy with modal logic. To see this, note that (2) is equivalent to the following proposition: whatever has logical consequences that are uncertain for me is itself uncertain for me. But this is clearly false. If you present me with a remote theorem of number theory I may well be uncertain of it, but I will be certain of the axioms just the same. Putting this point in terms of epistemic possibility, the negation of the theorem may be epistemically possible for me, but the negation of the conjunction of the axioms will not be. The reason for this discrepancy, of course, is that I may not be certain that the axioms do entail the theorem. This suggests that we modify premise (2) as follows:

- (2) If (P entails Q) is certain for me, and P is epistemically possible for me, then Q is epistemically possible for me.

In other words, epistemic possibility is transmitted by entailment *when I am certain that the entailment holds*.

The revised premise is undoubtedly true, but it brings a new difficulty along with it. To make the resulting argument valid, we shall have also to modify premise (3), as follows:

- (3) I am certain that the following entailment holds: (*T* seems evident to me and the Demon Hypothesis is true) entails (*T* is false).

Now we *are* claiming certainty in one of our premises. So what becomes of my attempt to vindicate Descartes against Prichard and Wittgenstein?

The answer is that it still succeeds. Prichard and Wittgenstein thought that *grounds* for doubt must be certain. Here they were wrong and Descartes was right. Grounds for doubt need only be epistemically possible, as my revised argument still illustrates. But something must be conceded to Prichard and Wittgenstein nonetheless. Although a ground for doubt need not be certain, one must in such a case be certain about the *logical relation* in which the ground stands to the dubitandum. This is the moral of our having to incorporate (3') into the revised argument. We may sum things up thus: doubt presupposes that *something* is certain; so far Prichard and Wittgenstein were right.²⁷ But the *ground* on which one bases a doubt need not be certain; on this point Descartes was right.

I want to make one observation before moving on. Look at my revised skeptical argument and I think you will agree that the only challengeable premise is (1). I think you will also agree that (1) is false only if the Demon Hypothesis is not epistemically possible. Thus, you can challenge the argument only by challenging the epistemic possibility of the Demon Hypothesis. And if you do this, you are claiming to be *certain* that the Demon Hypothesis is false. To overthrow the demon of skepticism, you must take him boldly by the horns.

V

In this section I shall present what I think is the most promising solution to the problem of the Cartesian Circle. Not only is it the best solution for Descartes, but it also has applications to contemporary epistemology.

The key to the solution I advocate is a distinction similar to one Anthony Kenny draws between the following propositions:

- (a) For all *P*, if I clearly and distinctly perceive that *P*, then I cannot doubt that *P*.
(b) I cannot doubt that (for all *P*, if I clearly and distinctly perceive that *P*, then *P*).²⁸

The distinction I shall use is analogous but different. For Kenny's indubitability, which does not obviously entail either evidence or truth, I want to substitute *certainty*. This gives us the following pair of propositions:

- (A) For all *P*, if I clearly and distinctly perceive that *P*, then I am certain that *P*.
(B) I am certain that (for all *P*, if I clearly and distinctly perceive that *P*, then *P*).

The difference is that (A) says that whenever I clearly and distinctly perceive any proposition I will be certain of *it* (the proposition in question), whereas (B) says that I am certain of a *general principle* connecting clear and distinct perception with truth. Clearly, (A) could be true even though (B) were false. (B) requires that I have the concept of clear and distinct perception, but (A) does not. Moreover, even if I did have this concept, I might be uncertain about the general connection between clear and distinct perception and truth, yet certain of every proposition I did clearly and distinctly perceive.²⁹

Unfortunately, the same English sentence – “I am certain of the truth of clear and distinct perceptions” – may be used to express *either* (A) *or* (B). Perhaps because of this, the distinction is often missed. But it is crucial to the Cartesian enterprise. I shall briefly indicate how it enables us to make sense of two otherwise puzzling passages in Descartes, and then I shall show how it provides an escape from the Circle.

The first passage is the notorious fourth paragraph in the Third Meditation, where Descartes appears to oscillate inconsistently between saying, on the one hand, God or no God, I am certain of things when I clearly and distinctly perceive them, and, on the other hand, I can doubt even the truth of clear and distinct perceptions if I do not know there is a veracious God. The appearance of inconsistency is removed if we see Descartes as being uncertain not of particular propositions that he clearly and distinctly perceives, but only of the general connection between clear and distinct perception and truth. What he shows us in this paragraph is that at this stage in the *Meditations* (A) is true of him but (B) is not.³⁰

The second thing our distinction enables us to understand is the epistemic advantage Descartes claims over the atheist. He concedes that even the atheist can be certain that the three angles of a triangle are equal to two right angles *if* he is clearly and distinctly perceiving this at the time. (*A*) is thus true of atheist and Descartes alike. But if at a later time both men merely *remember* having a clear and distinct perception of that theorem, Descartes will still be certain of it, but the atheist will not. This is *not* because (as the Memory Gambit would have it) Descartes can trust his memory and the atheist cannot. It is rather because Descartes can be certain (after he has proved the veracity of God) that anything he once clearly and distinctly perceived is true, whereas the atheist cannot. So (*B*) is true of Descartes, but not of the atheist.³¹

Now let us see how our distinction enables us to break out of the Circle. The first thing to notice is that (*B*) need not be true at the beginning of the Cartesian enterprise. I do *not* have to be certain that all clear and distinct perceptions are true before I prove that God exists. Proposition (2), which describes the lower arc of the Circle, is false.

Although (*B*) is false at the outset, however, it does not follow that (*A*) is false at the outset. And if (*A*) is true at any time, then anything I perceive clearly and distinctly at that time will be something I am certain of. Clear and distinct perception will thus provide me with an initial stock of premises I know for certain to be true. According to Descartes, these premises will include *I think, if I think then I exist, and a cause must contain at least as much reality as its effect*, among others. From these first principles I can go on to prove other things, including the existence of God and, eventually, the principle that all clear and distinct perceptions are true.³²

Clear and distinct perception and proposition (*A*) play crucial roles in my proposal that must not be misunderstood. The fact that I clearly and distinctly perceive a proposition does not serve as a *ground* for accepting it. It is a *source* of knowledge, but not a ground. Nor does proposition (*A*) serve as a ground. Rather, it is a *fact* that enables knowledge to get started. (We can authenticate this fact later if we wish, but need not do so in the beginning.)

Prichard misunderstands Descartes in just the way I am warning against. According to him, Descartes thinks he can arrive at certainty about a proposition *P* only by running through an argument of the following sort:

Whatever I perceive clearly and distinctly is certain.

I am perceiving clearly and distinctly that *P*.

Therefore, *P* is certain.³³

Here proposition (*A*) and my clearly and distinctly perceiving *P* both appear as grounds. As grounds, they cannot contribute to knowledge unless they are themselves known. But the question how I know *them* obviously has no satisfactory answer if an argument like the one above must stand behind any answer I give. Prichard's interpretation of Descartes thus leads to disaster (as he himself is quick to point out).

In opposition to Prichard's interpretation, I maintain that in order to become certain of a proposition I do not need to know that I am clearly and distinctly perceiving it, nor that whatever I so perceive is either certain or true. It is enough that I *do* clearly and distinctly perceive the proposition. (*A*) *says* that this is enough. For (*A*) says that perceiving something clearly and distinctly is *sufficient* to render me certain of it. It follows that nothing else is *necessary*, unless it is also necessary for the occurrence of clear and distinct perception in the first place. But neither knowledge of (*A*) nor knowledge of the fact that I am clearly and distinctly perceiving something is necessary for such perception to occur.³⁴

The point I have been insisting upon could be summed up as follows: (*A*) is not a principle I have to *apply* in order to gain knowledge; I need only *fall under* it.

The solution I am proposing does not require us to deny (1), the proposition describing the upper arc of the Circle. We can side with Descartes if we like and hold that in order to know that whatever is clearly and distinctly perceived is true, we must first know that God exists.³⁵ But this is not to say that before God's existence is known, clear and distinct perception affords us no certainty. On the contrary, it does, and this is what lets knowledge get started. Those who observe that we must be certain of some clear and distinct perceptions independently of our knowledge of God are therefore correct. But to conclude from this that (1) must be denied would be to confuse (*A*) with (*B*).

Moreover, to say (as I do) that we must be certain at the outset of some clear and distinct perceptions is not to say (as I don't) that we must be certain at the outset of the proposition *some clear and distinct perceptions are true*. Nor is it to say that we must be certain of a more specific

proposition of the form *those clear and distinct perceptions that are F* (for example, bathed in the light of nature) *are true*. In either of these cases we would be entertaining a proposition *about* the epistemic powers of clear and distinct perception, but no such thing is required of us.

We have seen that in virtue of (*not* by appeal to) proposition (*A*) we acquire an initial stock of certainties. Where do we go from here? Descartes's route is all too familiar. Among the initial certainties are premises that entail the existence of God. In the Third Meditation he clearly and distinctly perceives this entailment, thus becoming certain that God exists. In the Fourth Meditation he clearly and distinctly perceives, and thus becomes certain, that God could not be a deceiver. He goes on to infer (and become certain) that whatever he perceives clearly and distinctly is true.

It is important to recognize at this point that despite his usual formulation of it, Descartes's rule of clearness and distinctness (the C&D Rule, I shall call it) is not merely a rule of *truth*. It is also a rule of *evidence* and, indeed, of *certainty*. Many passages make this clear. When Descartes first introduces the C&D Rule, it is by means of the question, "Do I not . . . know what is requisite to render me certain of a truth?"³⁶ A few sentences later he refers to clear and distinct perceptions as "matters in which I believe myself to have the best evidence."³⁷ Elsewhere he says that clear and distinct perceptions are "true and certain."³⁸ Moreover, he often tells us that when he clearly and distinctly perceives something, he cannot help but believe it.³⁹ What is clearly and distinctly perceived, then, is not only *true*, but also *maximally evident and believed*. It is therefore *known for certain*. The C&D Rule thus turns out to be equivalent to proposition (*A*).

Descartes's procedure could be summed up thus: by falling under proposition (*A*) (that is, the C&D Rule), he becomes certain of premises from which he eventually derives proposition (*A*) itself.⁴⁰ But since he does not have to *use* proposition (*A*) at any step along the way, there is no circle.

Once Descartes knows that proposition (*A*) is true, of what use is the information? Well, for one thing, it enables him to cast off the atheist's handicap: he can now be certain of things he merely remembers having clearly and distinctly perceived. But more importantly, it enables him to vindicate his starting point. It gives him an answer to the critic who says, "I grant that your procedure is not

circular, but I don't see how you can escape the charge of arbitrariness in your first premises. What is the justification for starting from just the premises you did?" After he has proved proposition (*A*) Descartes can give the following reply: "Those premises are things I knew for certain. The proof of this is that I perceived them clearly and distinctly, and whatever I so perceive is certain."

To make the last point clearer, let me review the successive stages of the Cartesian enterprise. Here are the things of which Descartes is certain, listed in the order in which he becomes certain of them:

- | | | |
|-----|--|---|
| (1) | <i>I think</i> , the causal maxims, etc. | Propositions known because they are clearly and distinctly perceived. |
| (2) | <i>God exists</i> , <i>God is no deceiver</i> . | Propositions known because they are clearly and distinctly perceived to follow from premises at level (1). |
| (3) | <i>Whatever I perceive clearly and distinctly is certain</i> . | Principle known because it is clearly and distinctly perceived to follow from propositions at level (2). |
| (4) | <i>I perceive clearly and distinctly that I think</i> , etc. | New premises, one corresponding to each premise at level (1). ⁴¹ |
| (5) | <i>I am certain that I think</i> , etc. | Propositions known because they are clearly and distinctly perceived to follow from propositions at levels (3) and (4). |

The propositions at stage (1) are Descartes's first premises; he accepts them without any supporting grounds. This is what incurs the charge of arbitrariness (or dogmatism). But that charge may be answered as follows. If proposition (*A*) is true, then Descartes's first premises are *immediately* justified, that is, they are justified simply in virtue of being clearly and distinctly perceived, not because they inherit justification from other propositions. From these immediately justified beginnings Descartes goes on to derive (at level (3)) proposition (*A*) itself, which serves as a reason for the higher-order propositions (at level (5)) to the effect that his initial premises were justified (indeed, certain). This shows that the initial premises were not arbitrary.

The strategy I am imputing to Descartes is nicely described in a recent article by William P. Alston. He writes:

For any belief that one is immediately justified in believing, one *may* find adequate reasons for

accepting the proposition that one is so justified. The curse (of dogmatism) is taken off immediate justification at the lower level, just by virtue of the fact that propositions at the higher level are acceptable only on the basis of reasons.⁴²

This fits Descartes perfectly. His initial premises are immediately justified and certain, but the higher-order proposition that *says* they are certain is justified by appeal to reasons.⁴³

I shall consider one more objection to Descartes's procedure. This one charges that proposition (*A*) is arbitrary because one could have chosen *any* rule of evidence and justified it by the method I attribute to Descartes. My reply is twofold. First, proposition (*A*) is not arbitrary; it is the conclusion of an argument. One may wish to quarrel with the argument, but that would be a separate objection. Second, it is simply not true that *any* principle could be justified in the same manner as (*A*). It has at least to be a *true* principle, or else no first premises will be justified in virtue of it.

This completes my solution to the problem of the Cartesian Circle. Needless to say, I do not wish to endorse all the details of Descartes's reasoning. But I do maintain that in general outline it is sound, and in Part Two I shall use what is of value in it to throw light on contemporary issues.

Part Two

VI

In contemporary epistemology there is a movement away from foundationalist theories of justification toward coherentist theories. Wilfrid Sellars and Keith Lehrer, two of the leaders of this movement, have criticized foundationalism by raising doubts about its ability to justify its own epistemic principles. In this Part I shall draw on what we have learned about Descartes in Part One to show how these criticisms may be met.

Two doctrines are essential to foundationalism: (*i*) there is a class of propositions – the “foundations” – that are self-evident or immediately justified; and (*ii*) every proposition that is justified is so at least partly in virtue of standing in certain relations to the foundations.⁴⁴ Certain other doctrines are often associated with foundationalism, but they are not entailed by (*i*) and (*ii*). For example, Descartes was a foundationalist who held (*iii*) that the foundational propositions must be not

only immediately justified, but also certain and indubitable, (*iv*) that they are limited in scope to simple necessary truths and propositions about one's own mental states, and (*v*) that the superstructure is related to the foundations by deductive relations exclusively. One can defend (*i*) and (*ii*) without being committed to (*iii*), (*iv*), and (*v*).

Foundationalists often set forth principles specifying the conditions under which propositions of various types are justified. Usually called *epistemic principles*, they fall into two groups: principles that tell us that propositions of certain types are justified independently of their logical relations to other propositions, and principles that tell us that if some propositions are *already* justified, then any propositions that stand in such-and-such relations to them are *also* justified. Principles of the first sort I call *generation principles*, since they are principles whereby justification is generated in the first place, and those of the second sort I call *transmission principles*, since they are principles whereby justification is transmitted from some propositions to others. Generation principles are used to lay the foundations, transmission principles to erect the superstructure.

The general form of an epistemic principle is “If . . . then *P* is justified for *S*.” The antecedent of a sentence expressing a transmission principle will contain terms of epistemic appraisal (such as “evident,” “certain,” and so on), since it must mention other propositions and specify their epistemic status. But the antecedent of a sentence expressing a generation principle will *not* contain terms of epistemic appraisal.⁴⁵

Descartes's C&D Rule is a generation principle. It tells us that if someone is clearly and distinctly perceiving a proposition – a state we can describe without using epistemic terms – then that proposition is certain for him. Another example of a generation principle is Chisholm's Principle (*A*), which says that if a subject is in any of a designated group of “self-presenting states,” then it is evident to him that he is in the state in question.⁴⁶

An obvious example of a transmission principle is the principle that deduction transmits justification – more precisely, if *P* is justified for *S*, and the proposition that *P* entails *Q* is true and justified for *S*, then *Q* is justified for *S*. This is the *only* transmission principle Descartes allowed.⁴⁷ But most foundationalists countenance several others, including, perhaps, principles whereby justification is transmitted to propositions about the

physical world and propositions about the past. The best known list of such nondeductive transmission principles is Chisholm's.⁴⁸

Critics of foundationalism look upon its epistemic principles with suspicious eyes. "What is the justification for *them*?" they ask. Sellars and Lehrer contend that foundationalists have no satisfactory answer to this question, and urge us to adopt a coherentist view instead. Coherentism denies (i) and (ii), maintaining that a proposition may be justified in virtue of belonging to a coherent system of propositions none of which is immediately justified.

It will be convenient to sum up the Sellars–Lehrer critique in three theses:⁴⁹

- (I) Epistemic principles must themselves be known (or justified) if knowledge (or justified belief) is to arise in accordance with them.⁵⁰
- (II) There is no way to justify epistemic principles within a foundationalist framework.
- (III) There is a way to justify epistemic principles within a coherentist framework.

If (I) and (II) are both true, foundationalism leads to skepticism. If (III) is true, coherentism does not. So the upshot of the Sellars–Lehrer critique is this: if we wish to avoid skepticism, we must reject foundationalism in favor of coherentism. In the remaining sections, I shall explore possible foundationalist responses to this challenge, concentrating on (I) and (II).

VII

Premise (I) of the Sellars–Lehrer critique is widely taken for granted, but acceptance of it rests on a misunderstanding of how epistemic principles function. I have already discussed this misunderstanding in Part One. There we saw that Descartes did not need to know that the C&D Rule was true in order for clear and distinct perception to give him knowledge. For just the same reason, we can say in general that a subject need not know that an epistemic principle is true in order for the circumstance mentioned in its antecedent to give him knowledge.

The argument for this is very simple. An epistemic principle has the form "If . . . then *P* is justified for *S*." In other words, it says that the obtaining of whatever condition is specified in its antecedent is *sufficient* for *P*'s being justified for

S.⁵¹ Now it is a logical truth that if *X* is sufficient for *Y*, then there is no other condition *Z* that is *necessary* for *Y*, unless *Z* is also necessary for *X*. But knowledge of an epistemic principle is not necessary for the obtaining of its antecedent. Therefore, knowledge of an epistemic principle is not necessary for knowledge to arise in accordance with it. The first premise of the Sellars–Lehrer critique is false.

VIII

The argument that shows that knowledge of epistemic principles is not required also shows that knowledge of their antecedents is not required. This undermines a criticism Sellars levels against Chisholm's Principle (A). That principle, recall, says that if a subject is in any of a designated group of psychological states, then it is evident to him that he is in whatever state it is. Sellars says that this principle "seems to point to" arguments of the following form:

It is a fact that I am *F*.

So, it is reasonable to believe that I am *F*.⁵²

He then observes, "In order for any such argument to do the job, its premise would have to have authority, it would have to be something which it is reasonable to believe."⁵³ Presumably the difficulty with this is that if the conclusion were in question, the argument could not be expected to put it out of question.

To raise this objection is to misunderstand Chisholm in just the way that Prichard misunderstood Descartes. Chisholm's Principle is not supposed to function as a suppressed major premise under which a subject must subsume himself. The point is rather that just as having a clear and distinct perception of something puts one in a condition of knowing it, so being in a self-presenting state puts one in a condition in which it is evident to one that he is in that state.⁵⁴

There is a standard objection to coherentism that can be answered in the same way. Coherentists say that a proposition is justified if it coheres with a system of propositions of the right kind – for example, those already accepted by the subject, or by "the scientists of our culture circle," and so on. Foundationalists have been wont to object by asking, "How do you know *which* propositions belong to the system? And how do you know that

a given proposition *does* cohere with them?” But if the coherentist principle is *true* (and one may, of course, wish to question *this*) one need not know those things.

IX

In section VII we saw that epistemic principles need not be known in order for knowledge to arise in accordance with them. This does not mean, however, that the question “What justifies your principles?” is one that the foundationalist can brush aside. There are two reasons for this. In the first place, although some knowledge arises even if the principles are not known, there may be other knowledge that *does* depend on knowing the principles. In Descartes’s system, for example, one must have knowledge of the C&D Rule in order to obtain knowledge of propositions clearly and distinctly perceived at an earlier time. One must also have knowledge of the Rule in order to arrive at *higher-order* knowledge – knowledge that one knows. In the second place, even if knowledge of epistemic principles is not required for either of the two purposes just mentioned, we may seek it for its own sake when we embark upon epistemological inquiry. If it turns out that epistemic principles are justified at all, foundationalism will owe us an account of *how* they are justified. Otherwise it will not be a *complete* theory of justification. I turn, therefore, to examination of premise (II) of the Sellars–Lehrer critique.

If epistemic principles are justified within a foundationalist framework, this must be in one of two basic ways: either they are immediately justified, lying *at* the foundation, or they are mediately justified, resting *upon* the foundation. Within the latter alternative there are two subalternatives, which are best explained with the help of a term borrowed from Alston: an *epistemic proposition* is a singular proposition that attributes evidence, certainty, or some other epistemic characteristic to another proposition.⁵⁵ Now on the first subalternative, epistemic principles are justified *after* epistemic propositions, thus occupying a higher story in the edifice of knowledge, and on the second subalternative epistemic principles are justified *before* epistemic propositions, thus occupying a lower story in the edifice of knowledge.⁵⁶ This gives us three alternatives in all: the justification of epistemic principles may be immediate, it may be mediate and posterior to that of epistemic propo-

sitions, or it may be mediate and prior to that of epistemic propositions. I shall discuss all three alternatives, beginning with the last.

X

The first pattern to be discussed – mediate justification of epistemic principles prior to the justification of epistemic propositions – follows this general sequence:

- (1) Propositions known immediately.
- (2) Further propositions inferred from propositions at level (1).
- (3) Epistemic principles inferred from propositions at level (2).
- (4) Instantiations of antecedents of epistemic principles.
- (5) Epistemic propositions inferred from propositions at levels (3) and (4).

The reader who refers back to section V will see that this is precisely the pattern I attributed to Descartes. At level (1) we have the *cogito* and other things known immediately, at level (2) the existence and veracity of God, at level (3) the C&D Rule, at level (4) subsumptions of level (1) propositions under the Rule, and at level (5) epistemic propositions attributing certainty to the propositions at level (1).

I argued in section V that Descartes’s procedure is neither circular nor arbitrary. What is problematic about it is simply whether there are indeed valid inferences from level (1) to level (2) and from level (2) to level (3). As Hume observed, to have recourse to the veracity of God in order to prove the certainty of our perceptions is to make “a very unexpected circuit.”⁵⁷ Contemporary epistemologists who use Descartes’s pattern will no doubt want to avoid the circuit through theology, replacing his level (2) propositions by something more in keeping with a naturalistic world view. Let us therefore inquire into the prospects for a Cartesian epistemology naturalized.⁵⁸

Let us begin by taking a look at an attempt by Sellars to give a naturalistic derivation of Chisholm’s Principle (C), which reads as follows:

If there is a certain sensible characteristic *F* such that *S* believes that he perceives something to be *F*, then it is *evident* to *S* that he is perceiving something to have that characteristic *F*, and also that there is something that is *F*.⁵⁹

Sellars' derivation of this principle occurs as part of a larger coherentist strategy – in true Hegelian fashion, he is trying to concede the approximate truth of the foundationalist's principles while showing that they find their rationale only within his own coherentist system. Nonetheless, aspects of his derivation might be appropriated by foundationalists.

The idea, then, is to derive Principle (C) from naturalistic facts – facts of the sort establishable by scientific inquiry. What might these facts be? Sellars' candidates are facts about concept formation and language learning. Certain sentences – for example, “Here is a red apple” – are learned as directly conditioned responses to states of affairs obtaining in one's immediate vicinity. Now to say that a response *R* is conditioned to a stimulus *S* is to say that for the most part *R* occurs when and only when *S* occurs. Therefore, one who has learned how to use the sentence “Here is a red apple” will tend to utter it when and only when a red apple is present to his eyes. Hence, the utterance of that sentence by one who knows the language is a good indicator of the presence of a red apple. And one who finds himself uttering (or spontaneously inclined to utter) “Here is a red apple” can therefore take this utterance (or inclination) as evidence for the truth of the belief expressed by those words. We thus arrive (in Sellars' words) at “something very like” Chisholm's Principle (C).⁶⁰

But just *how* like Chisholm's Principle is it? What Sellars' argument really shows is at most this: if *S* utters (or is spontaneously inclined to utter) the words “Here is an *F*” (where “*F*” is any predicate directly conditioned to nonverbal stimuli), then the proposition expressed by those words is *likely to be true*. This differs from Principle (C) in several ways, one of which for our purposes is crucial: where (C) contains the term “evident” Sellars' principle contains the phrase “likely to be true.”⁶¹ It is plain that the latter must be taken in a *statistical* sense – the import of Sellars' principle is that most beliefs of a certain sort are true. Now the mere fact that most beliefs of a certain sort are true does not suffice to make those beliefs *evident* (or even justified to any degree) for the persons who hold them. After all, *any* true belief belongs to at least one class of beliefs most members of which are true, but not all true beliefs are justified. What our Sellarsian considerations establish, therefore, is a *statistical* principle, not an epistemic principle in the proper sense of the term.⁶²

This may account for Sellars' allegiance to premise (I). We saw in section VII that when epistemic principles are understood in the way I suggested, (I) is demonstrably false. But to someone who really has statistical rather than epistemic principles in mind, (I) will seem eminently plausible. The statistical fact that most beliefs of a certain sort are true does not make any of those beliefs evident; but if a believer *knew* this statistical fact, and knew in addition that one of his beliefs belonged to the sort in question, then (if no evidence pointed the other way) he *would* be justified in holding that belief. It does seem that statistical principles, unlike epistemic principles, can contribute to knowledge only if they are themselves known.⁶³

Whether this explains Sellars' acceptance of (I) or not, the point remains that his naturalistic derivation does not yield an epistemic principle, but only a statistical principle. Moreover, given a certain assumption that I have merely left tacit until now, I think it is safe to say that *any* attempt at a naturalistic derivation of epistemic principles would meet the same fate. The assumption is this: epistemic principles involve concepts that are *irreducibly* epistemic. That is to say, they involve concepts like evidence and certainty, and these concepts can only be defined with the help of *other* epistemic concepts; they cannot be defined solely in terms of logical and empirical concepts such as truth, probability, causation, and belief.⁶⁴ If this assumption is correct, it is hard to see how epistemic principles could ever be derived from propositions established by scientific inquiry. The difficulties involved would be analogous to those involved in trying to derive “ought”-statements from “is”-statements or observation statements from theoretical statements (without correspondence rules).

In conclusion, then, without a naturalistic reduction of epistemic concepts, there cannot be a naturalistic derivation of epistemic principles.⁶⁵ We must look elsewhere for their justification.⁶⁶

XI

The next pattern to be discussed – mediate justification of epistemic principles posterior to the justification of epistemic propositions – is also present, although not prominent, in the *Meditations*. At the beginning of the Third Meditation

Descartes introduces his C&D Rule in the following way:

I am certain that I am a thing which thinks; but then do I not likewise know what is requisite to render me certain of a truth? Certainly in this first knowledge there is nothing that assures me of its truth, excepting the clear and distinct perception of that which I state, which would not indeed suffice to assure me that what I say is true, if it could ever happen that a thing which I conceived so clearly and distinctly could be false; and accordingly it seems to me that already I can establish as a general rule that all things which I perceive very clearly and very distinctly are true.⁶⁷

Here Descartes starts with an epistemic proposition – *I am certain that I am a thing which thinks* – and moves to an epistemic principle – *all things which I perceive very clearly and distinctly are true*.⁶⁸

In the immediately following paragraphs, however, Descartes says that his principle is subject to doubts that can be removed only by establishing the existence of a veracious God. Evidently, then, he views the passage I have quoted as belonging merely to the context of discovery, not to the context of justification. In it the C&D Rule is first brought to light, but not yet established as known.

But why *not* regard the passage as a justification of the C&D Rule? We can extract from it the following argument:

- (1) I am certain that I am a thinking thing.
- (2) The only possible source of this certainty is clear and distinct perception.
- (3) Therefore, clear and distinct perception is a source of certainty (i.e., whatever I perceive clearly and distinctly is certain).

This argument fits the pattern Chisholm has labelled “critical cognitivism.” Such arguments contain one premise affirming that we do have knowledge, certainty, or justification with respect to propositions of a given type, another premise ruling out all possible sources of this knowledge except one, and a conclusion affirming that the remaining possible source must therefore be the source of the knowledge in question. Epistemic principles are thus justified by appeal to epistemic propositions.⁶⁹

Critical cognitivism arouses misgivings in many. I shall spend the rest of this section discussing the two strongest objections to it I have encountered. The first comes from Alston, who in the following passage is criticizing the view that epistemic propositions may be immediately justified, but whose remarks can be adapted against critical cognitivism as well:

In taking a belief to be justified, we are evaluating it in a certain way. And, like any evaluative property, epistemic justification is a supervenient property, the application of which is based on more fundamental properties. . . . Hence in order for me to be justified in believing that *S*'s belief that *p* is justified, I must be justified in certain other beliefs, viz., that *S*'s belief that *p* possesses a certain property, *Q* and that *Q* renders its possessor justified. (Another way of formulating this last belief is: a belief that there is a valid epistemic principle to the effect that any belief that is *Q* is justified.) Hence in no case can an epistemic belief that *S* is justified in believing that *p*, itself be immediately justified.⁷⁰

Alston is claiming that in order to come to be justified in believing an epistemic proposition, I must first be justified in believing an appropriate epistemic principle. If he is right about this, the justification of epistemic principles cannot be posterior to that of epistemic propositions, contrary to what is envisaged by critical cognitivism. Things would have to be the other way around.

Let us grant that justification is a supervenient property.⁷¹ There are still two lines of reply to Alston's argument. Using “*Jp*” to abbreviate “*S* is justified in believing that *p*,” we can symbolize his main premise as follows:

$$Jp \text{ requires } \exists Q \{JQp \ \& \ J\{p\}(Qp \rightarrow Jp)\}.$$

Now the first reply is this: why is it not sufficient for *Jp* to have

$$J\{\exists Q[Qp \ \& \ (p)(Qp \rightarrow Jp)]\}.$$

In other words, why is it not sufficient to be justified in believing that *p* possesses *some* property *Q* that renders its possessor justified without knowing *which* property it is? The job of the critical cognitivist's argument would then be to *identify* the property *Q*.

I find this reply quite plausible. Sometimes I reflect that I know a certain proposition, then ask myself *how* I know it (what makes me justified). I can be justified in my initial reflection even before I have successfully answered the question it provokes. (The analogue in ethics would be knowing that an act is right without yet knowing what *makes* it right.)

The second reply is this: if it is a true epistemic principle that (p) ($Qp \rightarrow Jp$), then a sufficient condition for Jp would be QJp , which does not itself require either JQp or $J(p) [Qp \rightarrow Jp]$. Hence, in order to defend his main premise, Alston needs the additional but unstated premise that no epistemic proposition can ever possess a property like Q . In short, he must rule out QJp . But why *can't* we have QJp ? Suppose that Q is the property of being clearly and distinctly perceived; why can't I clearly and distinctly perceive not only that $2 + 3 = 5$, but also that it is *evident* to me that $2 + 3 = 5$?⁷² Unless this alternative can be ruled out, Alston's argument is inconclusive.

The other objection to critical cognitivism is that it seems unsatisfyingly arbitrary and ad hoc. I do not know how to allay this misgiving, except by pointing out, as Chisholm does, that the conclusion of a critical cognitivist argument is (sometimes, at least) the consequence of premises that are individually quite plausible. For example, isn't it certain that I am (at least) a thinking thing? And what else could render me certain of it, if not the fact that I clearly and distinctly perceive it?

Before moving on I want to make three comments about the status of epistemic principles that are justified by critical cognitivist arguments. First, since knowledge of epistemic principles is not on this view a prerequisite for knowing that you know, there may be few occasions if any on which knowledge of them is called for. But this is not to say that they are idle; on the contrary, they have to be operative if there is to be knowledge at all. Second, every critical cognitivist argument has one premise that affirms that all possible sources but one of a given type of knowledge are barren. But it is hard to be sure that one has considered all possible sources, and harder still to be sure that one has considered all possible ways in which knowledge might arise from a given source. Therefore, the justification of epistemic principles afforded by critical cognitivism is bound to be somewhat tentative and conjectural. Third, it is sometimes suggested that epistemic principles are

synthetic and a priori.⁷³ But a critical cognitivist should not say this, for the argument he offers makes essential use of the a posteriori premise that we do know various things. If epistemic principles are a priori for him in any sense at all, it can only be in the Kantian sense that they are necessary presuppositions of knowledge.

XII

The last possibility to be discussed is that epistemic principles are immediately justified. The historical Descartes did not countenance this possibility, but I can imagine another Descartes meditating to himself as follows: "By reflecting on my condition when I clearly and distinctly perceive that $2 + 3 = 5$, I can see that my condition is one of *knowing*. Moreover, I can see that *any* state of clear and distinct perception would have to be a state of knowing. There could no more be a state of clear and distinct perception that was not a state of knowing than there could be adjacent mountains that did not enclose a valley." For this Descartes, the C&D Rule is an immediately justified necessary truth. Other foundationalists have claimed a similar status for their epistemic principles.⁷⁴

Lehrer objects strenuously to the imputation of immediate justification to epistemic principles.⁷⁵ Such a maneuver, he says, has the following disadvantages: (i) it makes the choice of epistemic principles "arbitrary";⁷⁶ (ii) it "entirely begs the question in favor of the foundation theorist" and "lacks dialectical cogency";⁷⁷ and (iii) it "opens the door to the most rampant forms of speculation."⁷⁸ What can the foundationalist say in reply?

(i) To say that the choice of principles is "arbitrary" is to say that there is no justification for choosing one set of principles rather than another, and Lehrer's ground for saying this is presumably that the foundationalist offers no reason in support of his choice. But if *this* is his ground, then Lehrer is presupposing that all justification is mediate – that nothing can be justified unless there is some further proposition that supports it. This, of course, begs the question *against* the foundation theorist.

(ii) In laying down a generation principle, a foundationalist affirms that the propositions specified in its consequent are immediately justified whenever its antecedent is true. If a coherentist, who denies that *any* propositions are immediately

justified, asks why he should accept such a principle, he will naturally be unsatisfied if he is told that it is itself immediately justified. So Lehrer is right: the foundationalist *does* beg the question, and his reply is dialectically ineffective.

But let us see if the coherentist can argue with greater dialectical effectiveness. He, too, espouses an epistemic principle, one that takes the following form: if a proposition coheres with a system of propositions of kind *K*, then it is justified.⁷⁹ Why should we accept a principle of *this* type? The coherentist dare not say that it is self-evident, lest he make a fundamental concession to the other side. It is more likely that he will say that it is justified by *its* coherence – in which case he is as guilty of begging the question as the foundationalist.

I mention this not to indulge in a *tu quoque*, but in order to bring out the fact that in the foundationalism-coherentism dispute, as in most matters of fundamental disagreement, it may be impossible for either side to support its view without begging the question against the other. It is hard for either side to get a dialectical grip on the other. One should not draw skeptical or relativist conclusions from this, however. One side may be in the right – and *know* that it is – even if it is incapable of demonstrating this to the other.

(iii) Let us consider, finally, the objection that foundationalism opens the door to speculation. If the foundationalist claims that his principles are immediately justified, then what is to prevent, let us say, a revelationist from claiming the same status for a principle to the effect that if *S* has an ostensible revelation that *P*, then *S* is justified in believing that *P*? The answer is that there is nothing to prevent this, but the foundationalist need not agree that the revelationist's principle is justified. Some claims to immediate justification are spurious.

Now let us see how speculation fares within the coherence theory. It is a consequence of this theory that *any* belief might be justified just so long as the believer enlists an appropriate cast of supporting beliefs.⁸⁰ Suppose, for instance, that someone entertains the notion that the world is made of cottage cheese, and that the rest of his beliefs form a system of kind *K* including the following items: the cosmos flowed forth from the teats of a

cow and curdled; when I sink my spade into the earth, up come mounds of creamy white stuff; and so on. What is to prevent *this*? Again, nothing. Moreover, there is an important difference between this situation and the one described in the last paragraph: the foundationalist did not have to agree that the revelationist's beliefs were justified; but the coherentist *does* have to agree that the beliefs of our eccentric cosmologist are justified (for him, at least), because they do, after all, satisfy the antecedent of the coherence principle. If anything, then, coherentism opens the door to speculation *wider*.

I conclude that there is nothing inherently objectionable about claiming that epistemic principles are immediately justified. Of course, this status can be claimed more plausibly for some principles than others. Perhaps few would hesitate to classify as immediately justified the principle that deduction transmits justification. And I would not hesitate to add to the list two generation principles: to cover the first truths a posteriori, Chisholm's Principle (*A*), and to cover the first truths a priori, a version of Descartes's C&D Rule, modified so as to be a rule of prima facie justification only. But to classify nondeductive transmission principles as immediately justified does not come as easily.

XIII

The results of Part Two may be summarized as follows. Coherentists have objected that epistemic principles must be justified, but cannot be if foundationalism is true. I have argued for just the opposite: epistemic principles need *not* be justified, but *can* be if foundationalism is true. I considered three ways in which this might occur. The first, naturalism, is perhaps the most exciting, but also, alas, the least promising. It overlooks the distinctively *epistemic* dimension of epistemic principles. The third, immediate justification, incurs no objection in principle, but very few epistemic principles can be claimed to be justified in this way. The second, critical cognitivism, is in some ways unsatisfying, but it, too, incurs no objection in principle, and may be the only alternative for nondeductive transmission principles.⁸¹

Notes

- 1 In presenting the problem this way I follow Willis Doney, "The Cartesian Circle," *Journal of the History of Ideas* 16 (1955), pp. 324–38. Cf. Arnauld's "only remaining scruple" in Haldane and Ross (eds), *Descartes: Philosophical Writings* (Cambridge: Cambridge University Press, 1911), vol. II, p. 92. (Hereafter I shall refer to the Haldane and Ross volumes as HR I and HR II.)
The more generalized form would be this: how can I know any epistemic principles unless I first know some other propositions from which to derive them? But how can I know those other propositions unless I first know some epistemic principles? See R. M. Chisholm, *The Problem of the Criterion*, The Aquinas Lecture, 1973 (Milwaukee: Marquette University Press, 1973). I shall not address myself to this problem explicitly, but what I would say about it will become clear in Part Two.
- 2 Doney, "The Cartesian Circle." Passages often cited in this connection are HR I, p. 184; HR II, pp. 38, 39, 114–15, and 245.
- 3 Throughout this paper I use "remember" in the sense of "ostensibly remember" so as not to ensure by definition that what is remembered is true.
- 4 Perhaps Doney does not wish to say that God guarantees that *whatever* we remember is true, but only that whatever we remember having perceived clearly and distinctly is something that we *did* perceive clearly and distinctly.
- 5 Harry Frankfurt, "Memory and the Cartesian Circle," *The Philosophical Review*, 71 (1962), 504–11.
- 6 This point is well made by A. K. Stout in "The Basis of Knowledge in Descartes," *Mind*, 38 (1929), 330–42 and 458–72. The article is reprinted in Doney (ed.), *Descartes* (Garden City: New York: Anchor Books, 1967), pp. 169–91. See also Frankfurt, "Memory and the Cartesian Circle."
- 7 Alan Gewirth, "The Cartesian Circle," *The Philosophical Review* 50 (1941), 368–95. I shall also refer to two other articles by Gewirth: "The Cartesian Circle Reconsidered," *The Journal of Philosophy* 67 (1970), pp. 685–700, and "Descartes: Two Disputed Questions," *The Journal of Philosophy*, 68 (1971), pp. 288–96.
- 8 "The Cartesian Circle," p. 374. "Clear and distinct perceptions are so coercive in their effect upon the mind that the mind cannot help assenting to them as true at the time it has such perceptions," *ibid.*, p. 383.
- 9 *Ibid.*, pp. 378 and 394.
- 10 Throughout the remainder of this paper I shall use "clear and distinct perception" (singular) for the act or faculty of perceiving clearly and distinctly, and "clear and distinct perceptions" (plural) for the propositional objects of such perception. That the objects of clear and distinct perception are *always* propositional is well argued by Frankfurt in *Demons, Dreamers, and Madmen* (Indianapolis: The Bobbs-Merrill Company, Inc., 1970), ch. 12.
- 11 My sources for what follows are the first two Gewirth articles mentioned in note 7. The steps I list are all present in the first article, but they are more explicitly set forth in the second.
- 12 By a "reason to doubt" *P* Descartes does not mean a reason to *disbelieve P* (i.e., believe its negation), but a reason to *withhold assent* from *P* (provided you want to assent only to what is certain).
- 13 Gewirth says that *R* need only *purport* to be clearly and distinctly perceived. But since he later disqualifies propositions from being reasons to doubt on the ground that they are not *in fact* clearly and distinctly perceived, he needs the stronger premise here.
- 14 Gewirth says that the existence and veracity of God is the *first* metaphysically certain proposition ("The Cartesian Circle," p. 394). But in fact the logic of his argument requires that all things we were previously psychologically certain of (because we perceived them clearly and distinctly) become metaphysically certain at one stroke the moment that proposition is established.
- 15 I pass over a third: why should the hypothesis of a deceiving God be the *only* reason for doubting clear and distinct perceptions?
- 16 HR I, p. 159; HR II, p. 277.
- 17 This point is made by Anthony Kenny in "The Cartesian Circle and the Eternal Truths," *The Journal of Philosophy* 67 (1970), pp. 685–700. In "Descartes: Two Disputed Questions," Gewirth makes the following reply: Descartes is entitled to the premise that clear and distinct perceptions are mutually consistent because consistency is an *internal matter* of relations among ideas, not an *external matter* of correspondence to fact. But this reply mistakenly assumes that Descartes's doubt is rooted in his representationalism, and arises only where there is a gap to be bridged between mental ideas and extramental facts. In fact his doubt is far more radical. To see this, notice three things: (i) the propositions we can clearly and distinctly perceive are limited to propositions about what is going on in our own minds and propositions about the relations among abstract entities; (ii) Descartes does *not* hold that mental happenings and abstract entities are known only via representatives – here there is no representationalist gap; yet (iii) he *still* finds room to wonder whether clear and distinct perception is a guarantee of truth.

- 18 Gewirth wants to say both that the absence of psychologically compelling reasons to doubt is sufficient for metaphysical certainty and that metaphysical certainty is sufficient for truth. But he cannot have it both ways.
- 19 Fred Feldman, "Epistemic Appraisal and the Cartesian Circle," *Philosophical Studies* 27 (1975), pp. 37–55.
- 20 I have recast his argument somewhat, but every premise I attribute to him is logically equivalent to a premise or definition he states in his article. I have also dropped references to persons and times.
- 21 Roderick Firth's "The Anatomy of Certainty," *The Philosophical Review* 76 (1967), pp. 3–27, divides senses of "certain" into three classes: truth-evaluating senses, warrant-evaluating senses, and testability-evaluating senses. In terms of this scheme, Descartes's sense of "certain" is both truth-evaluating and warrant-evaluating; Feldman's is warrant-evaluating but not truth-evaluating; and Gewirth's is neither warrant-evaluating nor truth-evaluating, but only (to extend Firth's classification) *belief-evaluating*.
- 22 Feldman considers the objection that a reason to doubt need only be a "metaphysical possibility," i.e., something of whose falsehood we are not metaphysically certain. (I prefer the term "epistemic possibility" here.) He gives both a textual and a strategic reply. The textual reply is that Descartes himself required that reasons to doubt be practically possible. But the passages Feldman cites in support of this are both inconclusive in themselves and outweighed by other passages, such as HR II, p. 266, where Descartes says that the "very least ground of suspicion" may engender doubt. The strategic reply is that by doing things his way "we help to provide a conceptual framework within which a solution to the problem at hand may be found." But I shall show that a solution may also be found within a framework that sets the standards for certainty higher.
- 23 H. A. Prichard, "Descartes's Meditations," in *Knowledge and Perception* (Oxford: Clarendon Press, 1950), pp. 71–104; reprinted in Doney (ed.), *Descartes*, pp. 140–68.
- 24 Ludwig Wittgenstein, *On Certainty* (New York: Harper & Row, 1969), sec. 115.
- 25 HR II, p. 273.
- 26 HR II, p. 277.
- 27 A corollary they draw is that universal doubt is impossible, but one should not make too much of this. Although one cannot have reasons for doubting everything, one could have reasons for doubting almost any arbitrary proposition. Though not universal skepticism, this is skepticism enough.
- 28 This distinction is anticipated in Kenny's book *Descartes* (New York: Random House, 1968) by the distinction on pp. 183–4 between first-order doubt (which would correspond to the denial of (a)) and

second-order doubt (which would correspond to the denial of (b)). It is drawn explicitly in his article "The Cartesian Circle and the Eternal Truths," cited previously.

One claim Kenny makes in connection with the distinction is incorrect. In the book he says that second-order doubt is doubt of particular propositions "in a roundabout manner" by referring to them "under some general heading, such as 'what seems to me most obvious'" (pp. 183–4). In the article he says that the doubt implied by the denial of (b) touches particular clearly and distinctly perceived propositions "only through referentially opaque wrappers" (p. 689). But referential opacity prevents any "touching" at all. One who doubts that what he perceives clearly and distinctly is true is *not* doubting particular clearly and distinctly perceived propositions in any manner, however roundabout, nor through any wrappers, however opaque. What he is doubting is whether there is any connection between clear and distinct perception and truth.

- 29 It seems also to be the case that (B) could be true even though (A) were false. I might be certain that clear and distinct perception guarantees truth, yet not certain that a given proposition I was clearly and distinctly perceiving was true, provided I was unaware that I *was* clearly and distinctly perceiving it.
- 30 This analysis of the paragraph is essentially Kenny's; see p. 689 of "The Cartesian Circle and the Eternal Truths." It must be said, however, that the final sentence of this paragraph – "Without a knowledge of these two truths [God exists and is not a deceiver] I do not see that I can ever be certain of anything" – is an embarrassment for almost any interpretation of Descartes. Here he digs himself into a pit so deep there can be no climbing out. Some interpreters – e.g., George Nakhnikian in "The Cartesian Circle Revisited," *American Philosophical Quarterly* 4 (1967), pp. 251–5 – regard this sentence as an aberration.
- 31 Descartes's advantage is not as great as he thinks. Merely *remembering* that he once had a clear and distinct perception of God's veracity will not (on pain of circularity) assure him *now* that anything he previously clearly and distinctly perceived is true. So Descartes must acquire a *present* clear and distinct perception of God's veracity by going through the theological proofs afresh. This means that he has an advantage over the atheist only in those cases where the theological proofs are shorter or more easily called to mind than the geometrical proofs the atheist must reconstruct.

It could be replied that the content of Descartes's recollection is not "I once clearly and distinctly perceived that God exists and is no deceiver," but just "God exists and is no deceiver." But if so, the atheist is entitled to say that the content of *his* recollection is not "I once clearly and distinctly

- perceived that the angles of a triangle are equal to two right angles," but just "The angles of a triangle are equal to two right angles."
- 32 Of course, I do not really believe that Descartes knew for certain everything he said he did, e.g., his causal maxims. But this does not detract from the soundness of the general plan I am attributing to him.
- 33 Doney (ed.), *Descartes*, p. 145.
- 34 Kenny hits upon an important part of the solution I am advocating when on p. 194 of *Descartes* he writes, "When [Descartes] passes from the clear and distinct perception of something to the affirmation of its truth, he does not do so by tacit appeal to a suppressed major premise; his affirmation is based directly on the intuition and not on a deduction derived from a general proposition about the truthfulness of intuitions." But three pages later he betrays this insight with the following claim: "If every other certainty is to be built upon the certainty afforded by clear and distinct perception, then it is essential, if there is to be any certainty at all of the type Descartes sought, that one should be able to be certain that one is clearly and distinctly perceiving something." On the contrary, if you don't need "Whatever is perceived clearly and distinctly is true" as a major premise, then you don't need "I am clearly and distinctly perceiving that *P*" as a minor premise.
- 35 Though in fact I would deny this, of course.
- 36 HR I, p. 158.
- 37 Ibid.
- 38 HR I, p. 184; HR II, p. 41. What Descartes says here is that clear and distinct perceptions are certain *after* God's existence is proved. For my purposes, we must interpret him to mean "From a knowledge of God, I can prove that whatever is perceived clearly and distinctly is certain" rather than "If I clearly and distinctly perceive something and I know that God exists, then I will be certain of that thing." The latter makes knowledge of God a prerequisite of certainty, but the former does not.
- 39 HR I, pp. 158, 176, 183; HR II, pp. 42, 266.
- 40 Once he becomes certain of (*A*), (*B*) is true, too – but not until then.
- 41 For a discussion of how premises of the form "I perceive clearly and distinctly that *P*" are known, see Frankfurt, *Demons, Dreamers, and Madmen*, ch. 13. Some critics of Descartes hold such knowledge to be problematic. It is an advantage of my interpretation that the need for it is postponed until level (4).
- 42 William P. Alston, "Two Types of Foundationalism," *The Journal of Philosophy* 73 (1976), pp. 165–85.
- 43 Alston himself does not attribute this strategy to Descartes. He thinks Descartes is an Iterative Foundationalist, i.e., one who believes that some propositions of the form "*P* is immediately justified for *S*" are themselves immediately justified for *S*. (See p. 182n.) But as I interpret Descartes, he is what Alston calls a Simple Foundationalist, i.e., one who believes that for any epistemic subject, *some* propositions are immediately justified, but that no propositions of the form "*P* is immediately justified for *S*" are immediately justified. Such higher-order propositions are justified only mediately by reference to the veracity of God.
- 44 By a self-evident proposition I do not mean one that derives its evidence from itself (whatever that might mean), but one that does not derive its evidence from any other propositions. (Similarly, when theologians speak of God as self-caused, they do not usually mean that God causes his own existence, but rather that his existence is not caused by any other being. See Caterus's remarks on this point at HR II, p. 4). Perhaps a better term is "immediately justified," which I shall generally use instead.
- It should be borne in mind that a justified proposition is not necessarily one that an epistemic subject has gone through the procedure of "justifying."
- 45 Some coherentists espouse principles that are neither generation principles nor transmission principles, but a sort of hybrid between the two. An example would be "If *P* coheres with the system of propositions accepted by *S* (or the scientists of our culture circle, etc.), then *P* is justified for *S*." This is like a generation principle in that its antecedent contains no epistemic terms, but like a transmission principle in that its antecedent specifies relations to other propositions. It seems to me, however, that an adequate epistemology must recognize at least one *full-blooded* generation principle. "Credibility may be transmitted from one statement to another through deductive or probability connections; but credibility does not spring from these connections by spontaneous generation," Nelson Goodman, "Sense and Certainty," *The Philosophical Review* 61 (1952), pp. 160–7.
- 46 Roderick M. Chisholm, *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1966), pp. 24–37 and 44.
- 47 The principle Descartes introduces in the Sixth Meditation – that beliefs in physical objects prompted by our sensory ideas must be true, else God would be a deceiver – is a rule of *truth*, not a rule of evidence. I say more about this distinction in Section X.
- 48 See Principles (*B*)–(*I*) in chapter III of *Theory of Knowledge*. Some readers may question my classification of Principles (*B*)–(*F*) as transmission principles. They have the form of generation principles, since their antecedents contain no terms of epistemic appraisal. I suspect, however, that Chisholm thinks of them as transmission principles. If they were generation principles, then he would be committed

- to the view that some propositions about the physical world and some propositions about the past are *immediately* justified; but Chisholm sides with the Cartesian tradition in holding that such propositions are only *mediately* justified.
- 49 Neither Sellars nor Lehrer presents exactly this argument, but the materials for it are contained in their writings. See Keith Lehrer, *Knowledge* (Oxford: Clarendon Press, 1974), pp. 143–4, and Wilfrid Sellars, “Givenness and Explanatory Coherence,” *The Journal of Philosophy* 70 (1973), 612–24, especially sections V–VII.
- 50 For convenience in what follows I shall often use the term “knowledge” where justification is all that need be at issue.
- 51 Sometimes epistemic principles are formulated so as to allow for the possibility that the justification arising in accordance with them may be defeated or overridden. In these cases the obtaining of the antecedent is not by itself sufficient for justification to arise; what is sufficient is this plus the absence of any overriding circumstances. But this complication does not affect the point I am making. (Incidentally, this same complication calls for a qualification in my earlier characterization of generation principles: a clause in the antecedent stipulating that there are no overriding circumstances *may*, in its specification of those circumstances, use terms of epistemic appraisal; but the *positive* clause in the antecedent will *not* use such terms.)
- 52 Sellars, “Epistemic Principles,” this vol., chap. 13; see section IV.
- 53 *Ibid.*
- 54 There are, of course, other objections that one could raise against Chisholm’s principle. Sellars’ most challenging one is this: a necessary condition of *P*’s being evident to anyone is his having learned a sentence that means that *P*; but being in one of the states Chisholm calls “self-presenting” is not sufficient for having learned any sentences; therefore, being in one of these states is not sufficient for anything’s being evident to you. Compare the argument on pp. 131–2 of “Empiricism and the Philosophy of Mind” in Sellars’ *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1963). The big issue raised by this objection, of course, and one too big to discuss here, is the relation of thought to language.
- 55 Alston, “Two Types of Foundationalism,” p. 169.
- 56 A third possibility, of course, is that epistemic principles and epistemic propositions occupy the *same* story of the edifice. But this is not very promising. There is some prospect of justifying epistemic principles if you can appeal to epistemic propositions, and some prospect of justifying epistemic propositions if you can appeal to epistemic principles, but little prospect of justifying each independently of the other. See Chisholm, “The Problem of the Criterion,” where the same point is made in different terms.
- 57 David Hume, *An Inquiry Concerning Human Understanding* (Indianapolis: Bobbs-Merrill, 1955), p. 162. Hume was referring to *sense* perceptions rather than clear and distinct perceptions, but the same point holds for the latter.
- 58 That naturalistic epistemology and Cartesian epistemology may be viewed as sharing the pattern (1)–(5) was suggested to me by Stephen Leeds.
- 59 *Theory of Knowledge*, p. 47.
- 60 This argument occurs both in “Epistemic Principles” and “Givenness and Explanatory Coherence.” For a somewhat similar account see W. V. Quine and Joseph Ullian, *The Web of Belief* (New York: Random House, 1970), pp. 33–6, where it is suggested that a “force for veracity” can be found “in the very mechanism of language learning.”
- 61 Other differences, not crucial for our purposes, are these: in Sellars’ principle the values of “*F*” are not restricted to sensible characteristics, and the talk of belief has been transposed into talk of utterance and inclination.
- 62 I classify as statistical any principle that is concerned with *frequency of truth* rather than *justification*. Thus, even a principle to the effect that *all* beliefs of a certain sort are true would count as statistical.
- 63 Sellars and Chisholm appear to agree in holding that statistical principles can be of epistemological significance only if they are themselves known. See *Science, Perception, and Reality*, pp. 167–8, and Chisholm, *Perceiving* (Ithaca: Cornell University Press, 1957), p. 27.
- 64 In “Empiricism and the Philosophy of Mind” Sellars seems to agree “The idea that epistemic facts can be analyzed without remainder – even ‘in principle’ – into nonepistemic facts . . . is, I believe, a radical mistake – a mistake of a piece with the so-called ‘naturalistic fallacy’ in ethics” – *Science, Perception, and Reality*, p. 131.
- 65 This point by itself, however, does not undermine naturalistic programs in epistemology. Naturalists might propose simply to *dispense* with epistemic principles, maintaining that statistical principles are all epistemology needs. To this it might be objected that knowledge of statistical principles presupposes the truth of at least one epistemic principle. After all, statistical principles must be inferred from data; surely we need a principle in virtue of which the data are justified and a principle in virtue of which the inference is justified. The naturalist can reply that this is true *only if knowledge is analyzed in terms of justification*. It is not true if knowledge is given an analysis from which the concept of justification is *eliminated*, as in Goldman’s “causal analysis” (Alvin Goldman, “A Causal Theory of Knowing,” *The Journal of Philosophy* 64 (1967), pp. 357–72) or Armstrong’s “reliability” analysis (David Arm-

- strong, *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press, 1973)). I do not believe that these analyses are adequate, but I lack the space to criticize them here.
- 66 Unfortunately, the criticisms raised in this section apply to Descartes's enterprise, too. Although it appeals to a supernatural entity, that enterprise is naturalistic in the sense that it tries to derive epistemic principles from nonepistemic facts.
- 67 HR I, p. 158.
- 68 Keep in mind that instead of "true" he should say "certain."
- 69 See *Theory of Knowledge*, pp. 59–61. There is a strain of critical cognitivism in Descartes's reply to the Second Objections. There he justifies the principle that "Every idea needs to have some really existing cause of its objective reality" by saying "The admission of this axiom is highly necessary for the reason that we must account for our knowledge of all things, both of sensuous and of nonsensuous objects, and do so by means of it alone" (HR II, p. 56).
- 70 Alston, "Two Types of Foundationalism," p. 170.
- 71 What I am granting is that justification is supervenient in the sense that its *instantiation* depends on that of nonepistemic properties. If Alston's point in calling justification supervenient is that its *warranted ascription* is always based on nonepistemic properties, I am challenging it.
- 72 Russell makes a similar suggestion on pp. 381–2 of *Human Knowledge* (New York: Simon and Schuster, 1948): "The degree of credibility attaching to a proposition is itself sometimes a datum. I think we should also hold that the degree of credibility to be attached to a *datum* is sometimes a datum."
- 73 E.g., by Chisholm in chapter 7 of *Perceiving*. But Chisholm also tends to favor critical cognitivism. Perhaps one could combine these positions by saying that epistemic principles are *discovered* through critical cognitivism, but *justified* by immediate intuition.
- 74 E.g., Chisholm in chapter 7 of *Perceiving*.
- 75 He also objects to the claim that epistemic principles are necessary truths, suggesting that a skeptic who denied them would not be contradicting himself. This is true if it means that a formal contradiction cannot be deduced from the skeptic's denial by appeal to logic and meanings alone. But not all necessary truths have denials that are self-contradictory in this sense. On the other hand, if (as he seems to be) Lehrer is using "self-contradictory" in the broad sense in which it is synonymous with "necessarily false," then it is no longer obvious that the skeptic's denial is *not* self-contradictory.
- 76 *Knowledge*, pp. 143–4.
- 77 *Ibid.*, p. 121.
- 78 *Ibid.*, pp. 152–3.
- 79 *Ibid.*, p. 54. Different ways of specifying what the coherence relation is and what a system of kind *K* is yield different coherentist principles, but the point I am making applies to them all.
- 80 And, in Lehrer's version, is a disinterested truth-seeker. See *ibid.*, pp. 189–90.
- 81 But there are other approaches I have not discussed. For example, according to the theory of meaning defended by John L. Pollock in *Knowledge and Justification* (Princeton, N.J.: Princeton University Press, 1974), epistemic principles would turn out to be true in virtue of the meanings of their constituent nonepistemic concepts.

Can Empirical Knowledge Have a Foundation?

Laurence Bonjour

The idea that empirical knowledge has, and must have, a *foundation* has been a common tenet of most major epistemologists, both past and present. There have been, as we shall see further below, many importantly different variants of this idea. But the common denominator among them, the central thesis of epistemological foundationism as I shall understand it here, is the claim that certain empirical beliefs possess a degree of epistemic justification or warrant which does not depend, inferentially or otherwise, on the justification of other empirical beliefs, but is instead somehow immediate or intrinsic. It is these non-inferentially justified beliefs, the unmoved (or self-moved) movers of the epistemic realm as Chisholm has called them,¹ that constitute the foundation upon which the rest of empirical knowledge is alleged to rest.

In recent years, the most familiar foundationist views have been subjected to severe and continuous attack. But this attack has rarely been aimed directly at the central foundationist thesis itself, and new versions of foundationism have been quick to emerge, often propounded by the erstwhile critics themselves. Thus foundationism has become a philosophical hydra, difficult to come to grips with and seemingly impossible to kill. The purposes of this paper are, first, to distinguish and clarify the main dialectical variants of foundationism, by viewing them as responses to one fundamental problem which is both the main motivation and the primary obstacle for foundationism; and second, as a result of this discussion to offer sche-

matic reasons for doubting whether any version of foundationism is finally acceptable.

The main reason for the impressive durability of foundationism is not any overwhelming plausibility attaching to the main foundationist thesis in itself, but rather the existence of one apparently decisive argument which seems to rule out all non-skeptical alternatives to foundationism, thereby showing that *some* version of foundationism must be true (on the assumption that skepticism is false). In a recent statement by Quinton, this argument runs as follows:

If any beliefs are to be justified at all, . . . there must be some terminal beliefs that do not owe their . . . credibility to others. For a belief to be justified it is not enough for it to be accepted, let alone merely entertained: there must also be good reason for accepting it. Furthermore, for an inferential belief to be justified the beliefs that support it must be justified themselves. There must, therefore, be a kind of belief that does not owe its justification to the support provided by others. Unless this were so no belief would be justified at all, for to justify any belief would require the antecedent justification of an infinite series of beliefs. The terminal . . . beliefs that are needed to bring the regress of justification to a stop need not be strictly self-evident in the sense that they somehow justify themselves. All that is required is that they should not owe their justification to any other beliefs.²

I shall call this argument *the epistemic regress argument*, and the problem which generates it, *the epistemic regress problem*. Since it is this argument

Originally published in *American Philosophical Quarterly* 16, 1 (1978), pp. 1-13.

which provides the primary rationale and argumentative support for foundationism, a careful examination of it will also constitute an exploration of the foundationist position itself. The main dialectical variants of foundationism can best be understood as differing attempts to solve the regress problem, and the most basic objection to the foundationist approach is that it is doubtful that any of these attempts can succeed. (In this paper, I shall be concerned with the epistemic regress argument and the epistemic regress problem only as they apply to empirical knowledge. It is obvious that an analogous problem arises also for *a priori* knowledge, but there it seems likely that the argument would take a different course. In particular, a foundationist approach might be incapable in an account of *a priori* knowledge.)

I

The epistemic regress problem arises directly out of the traditional conception of knowledge as *adequately justified true belief*³ – whether this be taken as a fully adequate definition of knowledge or, in light of the apparent counter-examples discovered by Gettier,⁴ as merely a necessary but not sufficient condition. (I shall assume throughout that the elements of the traditional conception are at least necessary for knowledge.) Now the most natural way to justify a belief is by producing a justificatory argument: belief *A* is justified by citing some other (perhaps conjunctive) belief *B*, from which *A* is inferable in some acceptable way and which is thus offered as a reason for accepting *A*.⁵ Call this *inferential justification*. It is clear, as Quinton points out in the passage quoted above, that for *A* to be genuinely justified by virtue of such a justificatory argument, *B* must itself be justified in some fashion; merely being inferable from an unsupported guess or hunch, e.g., would confer no genuine justification upon *A*.

Two further points about inferential justification, as understood here, must be briefly noted. First, the belief in question need not have been *arrived at* as the result of an inference in order to be inferentially justified. This is obvious, since a belief arrived at in some other way (e.g., as a result of wishful thinking) may later come to be maintained solely because it is now seen to be inferentially justifiable. Second, less obviously, a person for whom a belief is inferentially justified need not have explicitly rehearsed the justificatory argu-

ment in question to others or even to himself. It is enough that the inference be available to him if the belief is called into question by others or by himself (where such availability may itself be less than fully explicit) and that the availability of the inference be, in the final analysis, his reason for holding the belief.⁶ It seems clear that many beliefs which are quite sufficiently justified to satisfy the justification criterion for knowledge depend for their justification on inferences which have not been explicitly formulated and indeed which could not be explicitly formulated without considerable reflective effort (e.g., my current belief that this is the same piece of paper upon which I was typing yesterday).⁷

Suppose then that belief *A* is (putatively) justified via inference, thus raising the question of how the justifying premise-belief *B* is justified. Here again the answer may be in inferential terms: *B* may be (putatively) justified in virtue of being inferable from some further belief *C*. But then the same question arises about the justification of *C*, and so on, threatening an infinite and apparently vicious regress of epistemic justification. Each belief is justified only if an epistemically prior belief is justified, and that epistemically prior belief is justified only if a still prior belief is justified, etc., with the apparent result that justification can never get started – and hence that there is no justification and no knowledge. The foundationist claim is that only through the adoption of some version of foundationism can this skeptical consequence be avoided.

Prima facie, there seem to be only four basic possibilities with regard to the eventual outcome of this potential regress of epistemic justification: (i) the regress might terminate with beliefs for which no justification of any kind is available, even though they were earlier offered as justifying premises; (ii) the regress might proceed infinitely backwards with ever more new premise beliefs being introduced and then themselves requiring justification; (iii) the regress might circle back upon itself, so that at some point beliefs which appeared earlier in the sequence of justifying arguments are appealed to again as premises; (iv) the regress might terminate because beliefs are reached which are justified – unlike those in alternative (i) – but whose justification does not depend inferentially on other empirical beliefs and thus does not raise any further issue of justification with respect to such beliefs.⁸ The foundationist opts for the last alternative. His argument is that

the other three lead inexorably to the skeptical result, and that the second and third have additional fatal defects as well, so that some version of the fourth, foundationist alternative must be correct (assuming that skepticism is false).

With respect to alternative (i), it seems apparent that the foundationist is correct. If this alternative were correct, empirical knowledge would rest ultimately on beliefs which were, from an epistemic standpoint at least, entirely arbitrary and hence incapable of conferring any genuine justification. What about the other two alternatives?

The argument that alternative (ii) leads to a skeptical outcome has in effect already been sketched in the original formulation of the problem. One who opted for this alternative could hope to avoid skepticism only by claiming that the regress, though infinite, is not vicious; but there seems to be no plausible way to defend such a claim. Moreover, a defense of an infinite regress view as an account of how empirical knowledge is actually justified – as opposed to how it might in principle be justified – would have to involve the seemingly dubious thesis that an ordinary knower holds a literally infinite number of distinct beliefs. Thus it is not surprising that no important philosopher, with the rather uncertain exception of Peirce,⁹ seems to have advocated such a position.

Alternative (iii), the view that justification ultimately moves in a closed curve, has been historically more prominent, albeit often only as a dialectical foil for foundationism. At first glance, this alternative might seem even less attractive than the second. Although the problem of the knower having to have an infinite number of beliefs is no longer present, the regress itself, still infinite, now seems undeniably vicious. For the justification of each of the beliefs which figure in the circle seems now to presuppose *its own* epistemically prior justification: such a belief must, paradoxically, be justified before it can be justified. Advocates of views resembling alternative (iii) have generally tended to respond to this sort of objection by adopting a holistic conception of justification in which the justification of individual beliefs is subordinated to that of the closed systems of beliefs which such a view implies; the property of such systems usually appealed to as a basis for justification is internal *coherence*. Such coherence theories attempt to evade the regress problem by abandoning the view of justification as essentially involving a linear order of dependence (though a

non-linear view of justification has never been worked out in detail).¹⁰ Moreover, such a coherence theory of empirical knowledge is subject to a number of other familiar and seemingly decisive objections.¹¹ Thus alternative (iii) seems unacceptable, leaving only alternative (iv), the foundationist alternative, as apparently viable.

As thus formulated, the epistemic regress argument makes an undeniably persuasive case for foundationism. Like any argument by elimination, however, it cannot be conclusive until the surviving alternative has itself been carefully examined. The foundationist position may turn out to be subject to equally serious objections, thus forcing a re-examination of the other alternatives, a search for a further non-skeptical alternative, or conceivably the reluctant acceptance of the skeptical conclusion.¹² In particular, it is not clear on the basis of the argument thus far whether and how foundationism can itself solve the regress problem; and thus the possibility exists that the epistemic regress argument will prove to be a two-edged sword, as lethal to the foundationist as it is to his opponents.

II

The most straightforward interpretation of alternative (iv) leads directly to a view which I will here call *strong foundationism*. According to strong foundationism, the foundational beliefs which terminate the regress of justification possess sufficient epistemic warrant, independently of any appeal to inference from (or coherence with) other empirical beliefs, to satisfy the justification condition of knowledge and qualify as acceptable justifying premises for further beliefs. Since the justification of these *basic beliefs*, as they have come to be called, is thus allegedly not dependent on that of any other empirical belief, they are uniquely able to provide secure starting-points for the justification of empirical knowledge and stopping-points for the regress of justification.

The position just outlined is in fact a fairly modest version of strong foundationism. Strong foundationists have typically made considerably stronger claims on behalf of basic beliefs. Basic beliefs have been claimed not only to have sufficient non-inferential justification to qualify as knowledge, but also to be *certain*, *infallible*, *indubitable*, or *incorrigible* (terms which are usually not very carefully distinguished).¹³ And most of the

major attacks on foundationism have focused on these stronger claims. Thus it is important to point out that nothing about the basic strong foundationist response to the regress problem demands that basic beliefs be more than adequately justified. There might of course be other reasons for requiring that basic beliefs have some more exalted epistemic status or for thinking that in fact they do. There might even be some sort of indirect argument to show that such a status is a consequence of the sorts of epistemic properties which are directly required to solve the regress problem. But until such an argument is given (and it is doubtful that it can be), the question of whether basic beliefs are or can be certain, infallible, etc., will remain a relatively unimportant side-issue.

Indeed, many recent foundationists have felt that even the relatively modest version of strong foundationism outlined above is still too strong. Their alternative, still within the general aegis of the foundationist position, is a view which may be called *weak foundationism*. Weak foundationism accepts the central idea of foundationism – viz. that certain empirical beliefs possess a degree of independent epistemic justification or warrant which does not derive from inference or coherence relations. But the weak foundationist holds that these foundational beliefs have only a quite low degree of warrant, much lower than that attributed to them by even modest strong foundationism and insufficient by itself to satisfy the justification condition for knowledge or to qualify them as acceptable justifying premises for other beliefs. Thus this independent warrant must somehow be augmented if knowledge is to be achieved, and the usual appeal here is to coherence with other such minimally warranted beliefs. By combining such beliefs into larger and larger coherent systems, it is held, their initial, minimal degree of warrant can gradually be enhanced until knowledge is finally achieved. Thus weak foundationism, like the pure coherence theories mentioned above, abandons the linear conception of justification.¹⁴

Weak foundationism thus represents a kind of hybrid between strong foundationism and the coherence views discussed earlier, and it is often thought to embody the virtues of both and the vices of neither. Whether or not this is so in other respects, however, relative to the regress problem weak foundationism is finally open to the very same basic objection as strong foundationism, with essentially the same options available for meeting it. As we shall see, the key problem for

any version of foundationism is whether it can itself solve the regress problem which motivates its very existence, without resorting to essentially *ad hoc* stipulation. The distinction between the two main ways of meeting this challenge both cuts across and is more basic than that between strong and weak foundationism. This being so, it will suffice to concentrate here on strong foundationism, leaving the application of the discussion to weak foundationism largely implicit.

The fundamental concept of strong foundationism is obviously the concept of a basic belief. It is by appeal to this concept that the threat of an infinite regress is to be avoided and empirical knowledge given a secure foundation. But how can there be any empirical beliefs which are thus basic? In fact, though this has not always been noticed, the very idea of an epistemically basic empirical belief is extremely paradoxical. For on what basis is such a belief to be justified, once appeal to further empirical beliefs is ruled out? Chisholm's theological analogy, cited earlier, is most appropriate: a basic belief is in effect an epistemological unmoved (or self-moved) mover. It is able to confer justification on other beliefs, but apparently has no need to have justification conferred on it. But is such a status any easier to understand in epistemology than it is in theology? How can a belief impart epistemic "motion" to other beliefs unless it is itself in "motion"? And, even more paradoxically, how can a belief epistemically "move" itself?

This intuitive difficulty with the concept of a basic empirical belief may be elaborated and clarified by reflecting a bit on the concept of epistemic justification. The idea of justification is a generic one, admitting in principle of many specific varieties. Thus the acceptance of an empirical belief might be morally justified, i.e. justified as morally obligatory by reference to moral principles and standards; or pragmatically justified, i.e. justified by reference to the desirable practical consequences which will result from such acceptance; or religiously justified, i.e. justified by reference to specified religious texts or theological dogmas; etc. But none of these other varieties of justification can satisfy the justification condition for knowledge. Knowledge requires *epistemic* justification, and the distinguishing characteristic of this particular species of justification is, I submit, its essential or internal relationship to the cognitive goal of truth. Cognitive doings are epistemically justified, on this conception, only if and to the extent that

they are aimed at this goal – which means roughly that one accepts all and only beliefs which one has good reason to think are true.¹⁵ To accept a belief in the absence of such a reason, however appealing or even mandatory such acceptance might be from other standpoints, is to neglect the pursuit of truth; such acceptance is, one might say, *epistemically irresponsible*. My contention is that the idea of being epistemically responsible is the core of the concept of epistemic justification.¹⁶

A corollary of this conception of epistemic justification is that a satisfactory defense of a particular standard of epistemic justification must consist in showing it to be truth-conducive, i.e. in showing that accepting beliefs in accordance with its dictates is likely to lead to truth (and more likely than any proposed alternative). Without such a meta-justification, a proposed standard of epistemic justification lacks any underlying rationale. Why after all should an epistemically responsible inquirer prefer justified beliefs to unjustified ones, if not that the former are more likely to be true? To insist that a certain belief is epistemically justified, while confessing in the same breath that this fact about it provides no good reason to think that it is true, would be to render nugatory the whole concept of epistemic justification.

These general remarks about epistemic justification apply in full measure to any strong foundationist position and to its constituent account of basic beliefs. If basic beliefs are to provide a secure foundation for empirical knowledge, if inference from them is to be the sole basis for the justification of other empirical beliefs, then that feature, whatever it may be, in virtue of which a belief qualifies as basic must also constitute a good reason for thinking that the belief is true. If we let “ ϕ ” represent this feature, then for a belief B to qualify as basic in an acceptable foundationist account, the premises of the following justificatory argument must themselves be at least justified.¹⁷

- (i) Belief B has feature ϕ .
- (ii) Beliefs having feature ϕ are highly likely to be true.

Therefore, B is highly likely to be true.

Notice further that while either premise taken separately might turn out to be justifiable on an *a priori* basis (depending on the particular choice of ϕ), it seems clear that they could not both be thus justifiable. For B is *ex hypothesi* an empirical belief, and it is hard to see how a particular empirical

belief could be justified on a purely *a priori* basis.¹⁸ And if we now assume, reasonably enough, that for B to be justified for a particular person (at a particular time) it is necessary, not merely that a justification for B exist in the abstract, but that the person in question be in cognitive possession of that justification, we get the result that B is not basic after all since its justification depends on that of at least one other empirical belief. If this is correct, strong foundationism is untenable as a solution to the regress problem (and an analogous argument will show weak foundationism to be similarly untenable).

The foregoing argument is, no doubt, exceedingly obvious. But how is the strong foundationist to answer it? *Prima facie*, there seem to be only two general sorts of answer which are even remotely plausible, so long as the strong foundationist remains within the confines of the traditional conception of knowledge, avoids tacitly embracing skepticism, and does not attempt the heroic task of arguing that an empirical belief could be justified on a purely *a priori* basis. First, he might argue that although it is indeed necessary for a belief to be justified and *a fortiori* for it to be basic that a justifying argument of the sort schematized above be in principle available in the situation, it is *not* always necessary that the person for whom the belief is basic (or anyone else) know or even justifiably believe that it is available; instead, in the case of basic beliefs at least, it is sufficient that the premises for an argument of that general sort (or for some favored particular variety of such argument) merely be *true*, whether or not that person (or anyone else) justifiably believes that they are true. Second, he might grant that it is necessary both that such justification exist and that the person for whom the belief is basic be in cognitive possession of it, but insist that his cognitive grasp of the premises required for that justification does not involve further empirical beliefs which would then require justification, but instead involves cognitive states of a more rudimentary sort which do not themselves require justification: *intuitions* or *immediate apprehensions*. I will consider each of these alternatives in turn.

III

The philosopher who has come the closest to an explicit advocacy of the view that basic beliefs may be justified even though the person for whom they

are basic is not in any way in cognitive possession of the appropriate justifying argument is D. M. Armstrong. In his recent book, *Belief, Truth and Knowledge*,¹⁹ Armstrong presents a version of the epistemic regress problem (though one couched in terms of knowledge rather than justification) and defends what he calls an “Externalist” solution:

According to ‘Externalist’ accounts of non-inferential knowledge, what makes a true non-inferential belief a case of *knowledge* is some natural relation which holds between the belief-state...and the situation which makes the belief true. It is a matter of a certain relation holding between the believer and the world. (p. 157)

Armstrong’s own candidate for this “natural relation” is “that there must be a *law-like connection* between the state of affairs *Bap* [i.e. *a*’s believing that *p*] and the state of affairs that makes ‘*p*’ true such that, given *Bap*, it must be the case that *p*” (p. 166). A similar view seems to be implicit in Dretske’s account of perceptual knowledge in *Seeing and Knowing*, with the variation that Dretske requires for knowledge not only that the relation in question obtain, but also that the putative knower *believe* that it obtains – though *not* that this belief be justified.²⁰ In addition, it seems likely that various views of an ordinary-language stripe which appeal to facts about how language is learned either to justify basic belief or to support the claim that no justification is required would, if pushed, turn out to be positions of this general sort. Here I shall mainly confine myself to Armstrong, who is the only one of these philosophers who is explicitly concerned with the regress problem.

There is, however, some uncertainty as to how views of this sort in general and Armstrong’s view in particular are properly to be interpreted. On the one hand, Armstrong might be taken as offering an account of how basic beliefs (and perhaps others as well) satisfy the adequate-justification condition for knowledge; while on the other hand, he might be taken as simply repudiating the traditional conception of knowledge and the associated concept of epistemic justification, and offering a surrogate conception in its place – one which better accords with the “naturalistic” world-view which Armstrong prefers.²¹ But it is only when understood in the former way that externalism (to adopt Armstrong’s useful term) is of any immediate interest

here, since it is only on that interpretation that it constitutes a version of foundationism and offers a direct response to the anti-foundationist argument set out above. Thus I shall mainly focus on this interpretation of externalism, remarking only briefly at the end of the present section on the alternative one.

Understood in this way, the externalist solution to the regress problem is quite simple: the person who has a basic belief need not be in possession of any justified reason for his belief and indeed, except in Dretske’s version, need not even think that there is such a reason; the status of his belief as constituting knowledge (if true) depends solely on the external relation and not at all on his subjective view of the situation. Thus there are no further empirical beliefs in need of justification and no regress.

Now it is clear that such an externalist position succeeds in avoiding the regress problem and the anti-foundationist argument. What may well be doubted, however, is whether this avoidance deserves to be considered a *solution*, rather than an essentially *ad hoc* evasion, of the problem. Plainly the sort of “external” relation which Armstrong has in mind would, if known, provide a basis for a justifying argument along the lines sketched earlier, roughly as follows:

- (i) Belief *B* is an instance of kind *K*.
- (ii) Beliefs of kind *K* are connected in a law-like way with the sorts of states of affairs which would make them true, and therefore are highly likely to be true.

Therefore, *B* is highly likely to be true.

But precisely what generates the regress problem in the first place is the requirement that for a belief *B* to be epistemically justified for a given person *P*, it is necessary, not just that there be justifiable or even true premises available in the situation which could in principle provide a basis for a justification of *B*, but that *P* himself know or at least justifiably believe some such set of premises and thus be in a position to employ the corresponding argument. The externalist position seems to amount merely to waiving this general requirement in cases where the justification takes a certain form, and the question is why this should be acceptable in these cases when it is not acceptable generally. (If it were acceptable generally, then it would seem that any true belief would be justified for any person, and the distinction between knowledge and true belief would collapse.) Such a move seems rather

analogous to solving a regress of causes by simply stipulating that although most events must have a cause, events of a certain kind need not.

Whatever plausibility attaches to externalism seems to derive from the fact that if the external relation in question genuinely obtains, then *P* will not go wrong in accepting the belief, and it is, in a sense, not an accident that this is so. But it remains unclear how these facts are supposed to justify *P*'s acceptance of *B*. It is clear, of course, that an external observer who knew both that *P* accepted *B* and that there was a law-like connection between such acceptance and the truth of *B* would be in a position to construct an argument to justify *his own* acceptance of *B*. *P* could thus serve as a useful epistemic instrument, a kind of cognitive thermometer, for such an external observer (and in fact the example of a thermometer is exactly the analogy which Armstrong employs to illustrate the relationship which is supposed to obtain between the person who has the belief and the external state of affairs (p. 166ff.)). But *P* himself has no reason at all for thinking that *B* is likely to be true. From his perspective, it is an accident that the belief is true.²² And thus his acceptance of *B* is no more rational or responsible from an epistemic standpoint than would be the acceptance of a subjectively similar belief for which the external relation in question failed to obtain.²³

Nor does it seem to help matters to move from Armstrong's version of externalism, which requires only that the requisite relationship between the believer and the world obtain, to the superficially less radical version apparently held by Dretske, which requires that *P* also believe that the external relation obtains, but does not require that this latter belief be justified. This view may seem slightly less implausible, since it at least requires that the person have some idea, albeit unjustified, of why *B* is likely to be true. But this change is not enough to save externalism. One way to see this is to suppose that the person believes the requisite relation to obtain on some totally irrational and irrelevant basis, e.g. as a result of reading tea leaves or studying astrological charts. If *B* were an ordinary, non-basic belief, such a situation would surely preclude its being justified, and it is hard to see why the result should be any different for an allegedly basic belief.

Thus it finally seems possible to make sense of externalism only by construing the externalist as simply abandoning the traditional notion of epistemic justification and along with it anything

resembling the traditional conception of knowledge. (As already remarked, this may be precisely what the proponents of externalism intend to be doing, though most of them are not very clear on this point.) Thus consider Armstrong's final summation of his conception of knowledge:

Knowledge of the truth of particular matters of fact is a belief which must be true, where the 'must' is a matter of law-like necessity. Such knowledge is a reliable representation or 'mapping' of reality. (p. 220)

Nothing is said here of reasons or justification or evidence or having the right to be sure. Indeed the whole idea, central to the western epistemological tradition, of knowledge as essentially the product of reflective, critical, and rational inquiry has seemingly vanished without a trace. It is possible of course that such an altered conception of knowledge may be inescapable or even in some way desirable, but it constitutes a solution to the regress problem or any problem arising out of the traditional conception of knowledge only in the radical and relatively uninteresting sense that to reject that conception is also to reject the problems arising out of it. In this paper, I shall confine myself to less radical solutions.

IV

The externalist solution just discussed represents a very recent approach to the justification of basic beliefs. The second view to be considered is, in contrast, so venerable that it deserves to be called the standard foundationist solution to the problem in question. I refer of course to the traditional doctrine of cognitive givenness, which has played a central role in epistemological discussions at least since Descartes. In recent years, however, the concept of the given, like foundationism itself, has come under serious attack. One upshot of the resulting discussion has been a realization that there are many different notions of givenness, related to each other in complicated ways, which almost certainly do not stand or fall together. Thus it will be well to begin by formulating the precise notion of givenness which is relevant in the present context and distinguishing it from some related conceptions.

In the context of the epistemic regress problem, givenness amounts to the idea that basic beliefs are

justified by reference, not to further *beliefs*, but rather to states of affairs in the world which are “immediately apprehended” or “directly presented” or “intuited.” This justification by reference to non-cognitive states of affairs thus allegedly avoids the need for any further justification and thereby stops the regress. In a way, the basic gambit of givenism (as I shall call positions of this sort) thus resembles that of the externalist positions considered above. In both cases the justificatory appeal to further beliefs which generates the regress problem is avoided for basic beliefs by an appeal directly to the non-cognitive world; the crucial difference is that for the givenist, unlike the externalist, the justifying state of affairs in the world is allegedly apprehended *in some way* by the believer.

The givenist position to be considered here is significantly weaker than more familiar versions of the doctrine of givenness in at least two different respects. In the first place, the present version does not claim that the given (or, better, the apprehension thereof) is certain or even incorrigible. As discussed above, these stronger claims are inessential to the strong foundationalist solution to the regress problem. If they have any importance at all in this context it is only because, as we shall see, they might be thought to be entailed by the only very obvious intuitive picture of how the view is supposed to work. In the second place, givenism as understood here does not involve the usual stipulation that only one’s private mental and sensory states can be given. There may or may not be other reasons for thinking that this is in fact the case, but such a restriction is not part of the position itself. Thus both positions like that of C. I. Lewis, for whom the given is restricted to private states apprehended with certainty, and positions like that of Quinton, for whom ordinary physical states of affairs are given with no claim of certainty or incorrigibility being involved, will count as versions of givenism.

As already noted, the idea of givenness has been roundly criticized in recent philosophical discussion and widely dismissed as a piece of philosophical mythology. But much at least of this criticism has to do with the claim of certainty on behalf of the given or with the restriction to private, subjective states. And some of it at least has been mainly concerned with issues in the philosophy of mind which are only distantly related to our present epistemological concerns. Thus even if the objections offered are cogent against other

and stronger versions of givenness, it remains unclear whether and how they apply to the more modest version at issue here. The possibility suggests itself that modest givenness may not be a myth, even if more ambitious varieties are, a result which would give the epistemological foundationalist all he really needs, even though he has usually, in a spirit of philosophical greed, sought considerably more. In what follows, however, I shall sketch a line of argument which, if correct, will show that even modest givenism is an untenable position.²⁴

The argument to be developed depends on a problem within the givenist position which is surprisingly easy to overlook. I shall therefore proceed in the following way. I shall first state the problem in an initial way, then illustrate it by showing how it arises in one recent version of givenism, and finally consider whether any plausible solution is possible. (It will be useful for the purposes of this discussion to make two simplifying assumptions, without which the argument would be more complicated, but not essentially altered. First, I shall assume that the basic belief which is to be justified by reference to the given or immediately apprehended state of affairs is just the belief that this same state of affairs obtains. Second, I shall assume that the given or immediately apprehended state of affairs is not itself a belief or other cognitive state.)

Consider then an allegedly basic belief that-*p* which is supposed to be justified by reference to a given or immediately apprehended state of affairs that-*p*. Clearly what justifies the belief is not the state of affairs simpliciter, for to say that would be to return to a form of externalism. For the givenist, what justifies the belief is the *immediate apprehension* or *intuition* of the state of affairs. Thus we seem to have three items present in the situation: the belief, the state of affairs which is the object of the belief, and the intuition or immediate apprehension of that state of affairs. The problem to be raised revolves around the nature of the last of these items, the intuition or immediate apprehension (hereafter I will use mainly the former term). It *seems* to be a cognitive state, perhaps somehow of a more rudimentary sort than a belief, which involves the thesis or assertion that-*p*. Now if this is correct, it is easy enough to understand in a rough sort of way how an intuition can serve to justify a belief with this same assertive content. The problem is to understand why the intuition, involving as it does the cognitive thesis that-*p*,

does not *itself* require justification. And if the answer is offered that the intuition is justified by reference to the state of affairs that-*p*, then the question will be why this would not require a second intuition or other apprehension of the state of affairs to justify the original one. For otherwise one and the same cognitive state must somehow constitute both an apprehension of the state of affairs and a justification of that very apprehension, thus pulling itself up by its own cognitive bootstraps. One is reminded here of Chisholm's claim that certain cognitive states justify themselves,²⁵ but that extremely paradoxical remark hardly constitutes an explanation of how this is possible.

If, on the other hand, an intuition is not a cognitive state and thus involves no cognitive grasp of the state of affairs in question, then the need for a justification for the intuition is obviated, but at the serious cost of making it difficult to see how the intuition is supposed to justify the belief. If the person in question has no cognitive grasp of that state of affairs (or of any other) by virtue of having such an intuition, then how does the intuition give him a *reason* for thinking that his belief is true or likely to be true? We seem again to be back to an externalist position, which it was the whole point of the category of intuition or givenness to avoid.

As an illustration of this problem, consider Quinton's version of givenism, as outlined in his book *The Nature of Things*.²⁶ As noted above, basic beliefs may, according to Quinton, concern ordinary perceptible states of affairs and need not be certain or incorrigible. (Quinton uses the phrase "intuitive belief" as I have been using "basic belief" and calls the linguistic expression of an intuitive belief a "basic statement"; he also seems to pay very little attention to the difference between beliefs and statements, shifting freely back and forth between them, and I will generally follow him in this.) Thus "this book is red" might, in an appropriate context, be a basic statement expressing a basic or intuitive belief. But how are such basic statements (or the correlative beliefs) supposed to be justified? Here Quinton's account, beyond the insistence that they are not justified by reference to further beliefs, is seriously unclear. He says rather vaguely that the person is "aware" (p. 129) or "directly aware" (p. 139) of the appropriate state of affairs, or that he has "direct knowledge" (p. 126) of it, but he gives no real account of the nature or epistemological status of this state of "direct awareness" or "direct knowledge," though

it seems clear that it is supposed to be a cognitive state of some kind. (In particular, it is not clear what "direct" means, over and above "non-inferential.")²⁷

The difficulty with Quinton's account comes out most clearly in his discussion of its relation to the correspondence theory of truth:

The theory of basic statements is closely connected with the correspondence theory of truth. In its classical form that theory holds that to each true statement, whatever its form may be, a fact of the same form corresponds. The theory of basic statements indicates the point at which correspondence is established, at which the system of beliefs makes its justifying contact with the world. (p. 139)

And further on he remarks that the truth of basic statements "is directly determined by their correspondence with fact" (p. 143). (It is clear that "determined" here means "epistemically determined.") Now it is a familiar but still forceful idealist objection to the correspondence theory of truth that if the theory were correct we could never know whether any of our beliefs were true, since we have no perspective outside our system of beliefs from which to see that they do or do not correspond. Quinton, however, seems to suppose rather blithely that intuition or direct awareness provides just such a perspective, from which we can in some cases apprehend both beliefs and world and judge whether or not they correspond. And he further supposes that the issue of justification somehow does not arise for apprehensions made from this perspective, though without giving any account of how or why this is so.

My suggestion here is that no such account can be given. As indicated above, the givenist is caught in a fundamental dilemma: if his intuitions or immediate apprehensions are construed as cognitive, then they will be both capable of giving justification and in need of it themselves; if they are non-cognitive, then they do not need justification but are also apparently incapable of providing it. This, at bottom, is why epistemological givenness is a myth.²⁸

Once the problem is clearly realized, the only possible solution seems to be to split the difference by claiming that an intuition is a semi-cognitive or quasi-cognitive state,²⁹ which resembles a belief in its capacity to confer justification, while differing from a belief in not requiring justification itself. In

fact, some such conception seems to be implicit in most if not all givenist positions. But when stated thus baldly, this "solution" to the problem seems hopelessly contrived and *ad hoc*. If such a move is acceptable, one is inclined to expostulate, then once again any sort of regress could be solved in similar fashion. Simply postulate a final term in the regress which is sufficiently similar to the previous terms to satisfy, with respect to the penultimate term, the sort of need or impetus which originally generated the regress; but which is different enough from previous terms so as not itself to require satisfaction by a further term. Thus we would have semi-events, which could cause but need not be caused; semi-explanata, which could explain but need not be explained; and semi-beliefs, which could justify but need not be justified. The point is not that such a move is always incorrect (though I suspect that it is), but simply that the nature and possibility of such a convenient regress-stopper needs at the very least to be clearly and convincingly established and explained before it can constitute a satisfactory solution to any regress problem.

The main account which has usually been offered by givenists of such semi-cognitive states is well suggested by the terms in which immediate or intuitive apprehensions are described: "immediate," "direct," "presentation," etc. The underlying idea here is that of *confrontation*: in intuition, mind or consciousness is directly confronted with its object, without the intervention of any sort of intermediary. It is in this sense that the object is *given* to the mind. The root metaphor underlying this whole picture is vision: mind or consciousness is likened to an immaterial eye, and the object of intuitive awareness is that which is directly before the mental eye and open to its gaze. If this metaphor were to be taken seriously, it would become relatively simple to explain how there can be a cognitive state which can justify but does not require justification. (If the metaphor is to be taken seriously enough to do the foundationist any real good, it becomes plausible to hold that the intuitive cognitive states which result would after all have to be infallible. For if all need for justification is to be precluded, the envisaged relation of confrontation seemingly must be conceived as too intimate to allow any possibility of error. To the extent that this is so, the various arguments which have been offered against the notion of infallible cognitive states count also against this version of givenism.)

Unfortunately, however, it seems clear that the mental eye metaphor will not stand serious scrutiny. The mind, whatever else it may be, is not an eye or, so far as we know, anything like an eye. Ultimately the metaphor is just far too simple to be even minimally adequate to the complexity of mental phenomena and to the variety of conditions upon which such phenomena depend. This is not to deny that there is considerable intuitive appeal to the confrontational model, especially as applied to perceptual consciousness, but only to insist that this appeal is far too vague in its import to adequately support the very specific sorts of epistemological results which the strong foundationist needs. In particular, even if empirical knowledge at some point involves some sort of confrontation or seeming confrontation, this by itself provides no clear reason for attributing epistemic justification or reliability, let alone certainty, to the cognitive states, whatever they may be called, which result.

Moreover, quite apart from the vicissitudes of the mental eye metaphor, there are powerful independent reasons for thinking that the attempt to defend givenism by appeal to the idea of a semi-cognitive or quasi-cognitive state is fundamentally misguided. The basic idea, after all, is to distinguish two aspects of a cognitive state, its capacity to justify other states and its own need for justification, and then try to find a state which possesses only the former aspect and not the latter. But it seems clear on reflection that these two aspects cannot be separated, that it is one and the same feature of a cognitive state, viz. its assertive content, which both enables it to confer justification on other states and also requires that it be justified itself. If this is right, then it does no good to introduce semi-cognitive states in an attempt to justify basic beliefs, since to whatever extent such a state is capable of conferring justification, it will to that very same extent require justification. Thus even if such states do exist, they are of no help to the givenist in attempting to answer the objection at issue here.³⁰

Hence the givenist response to the anti-foundationist argument seems to fail. There seems to be no way to explain how a basic cognitive state, whether called a belief or an intuition, can be directly justified by the world without lapsing back into externalism – and from there into skepticism. I shall conclude with three further comments aimed at warding off certain likely sorts of misunderstanding. First. It is natural in this connection to attempt to justify basic beliefs by appealing

to *experience*. But there is a familiar ambiguity in the term “experience,” which in fact glosses over the crucial distinction upon which the foregoing argument rests. Thus “experience” may mean either an *experiencing* (i.e., a cognitive state) or something *experienced* (i.e., an object of cognition). And once this ambiguity is resolved, the concept of experience seems to be of no particular help to the givenist. Second. I have concentrated, for the sake of simplicity, on Quinton’s version of givenism in which ordinary physical states of affairs are among the things which are given. But the logic of the argument would be essentially the same if it were applied to a more traditional version like Lewis’s in which it is private experiences which are given, and I cannot see that the end result would be different – though it might be harder to discern, especially in cases where the allegedly basic belief is a belief about another cognitive state. Third. Notice carefully that the problem raised here with respect to givenism is a logical problem (in a broad sense of “logical”). Thus it would be a mistake to think that it can be solved simply by indicating some sort of state which seems intuitively to have the appropriate sorts of characteristics; the problem is to understand how it is *possible*

for any state to have those characteristics. (The mistake would be analogous to one occasionally made in connection with the free-will problem: the mistake of attempting to solve the logical problem of how an action can be not determined but also not merely random by indicating a subjective act of effort or similar state, which seems intuitively to satisfy such a description.)

Thus foundationism appears to be doomed by its own internal momentum. No account seems to be available of how an empirical belief can be genuinely justified in an epistemic sense, while avoiding all reference to further empirical beliefs or cognitions which themselves would require justification. How then is the epistemic regress problem to be solved? The natural direction to look for an answer is to the coherence theory of empirical knowledge and the associated non-linear conception of justification which were briefly mentioned above.³¹ But arguments by elimination are dangerous at best: there may be further alternatives which have not yet been formulated; and the possibility still threatens that the epistemic regress problem may in the end be of aid and comfort only to the skeptic.

Notes

- 1 Roderick M. Chisholm, *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1966), p. 30.
- 2 Anthony Quinton, *The Nature of Things* (London: Routledge and Kegan Paul, 1973), p. 119. This is an extremely venerable argument, which has played a central role in epistemological discussion at least since Aristotle’s statement of it in the *Posterior Analytics*, Book I, ch. 2–3. (Some have found an anticipation of the argument in the *Theaetetus* at 209E–210B, but Plato’s worry in that passage appears to be that the proposed definition of knowledge is circular, not that it leads to an infinite regress of justification.)
- 3 “Adequately justified” because a belief could be justified to some degree without being sufficiently justified to qualify as knowledge (if true). But it is far from clear just how much justification is needed for adequacy. Virtually all recent epistemologists agree that certainty is not required. But the lottery paradox shows that adequacy cannot be understood merely in terms of some specified level of probability. (For a useful account of the lottery paradox, see Robert Ackermann, *Knowledge and Belief* (Garden City, NY: Doubleday, 1972), pp. 39–50.) Armstrong, in *Belief, Truth and Knowledge* (London: Cambridge University Press, 1973), argues that what is required

is that one’s reasons for the belief be “conclusive,” but the precise meaning of this is less than clear. Ultimately, it may be that the concept of knowledge is simply too crude for refined epistemological discussion, so that it may be necessary to speak instead of degrees of belief and corresponding degrees of justification. I shall assume (perhaps controversially) that the proper solution to this problem will not affect the issues to be discussed here, and speak merely of the reasons or justification making the belief *highly likely* to be true, without trying to say exactly what this means.

- 4 See Edmund Gettier, “Is Justified True Belief Knowledge?” this vol., ch. 7. Also Ackermann, *Knowledge and Belief*, ch. V, and the corresponding references.
- 5 For simplicity, I will speak of inference relations as obtaining between beliefs rather than, more accurately, between the propositions which are believed. “Inference” is to be understood here in a very broad sense; any relation between two beliefs which allows one, if accepted, to serve as a good reason for accepting the other will count as inferential.
- 6 It is difficult to give precise criteria for when a given reason is *the* reason for a person’s holding a belief.

- G. Harman, in *Thought* (Princeton: Princeton University Press, 1973), argues that for a person to believe for a given reason is for that reason to *explain* why he holds that belief. But this suggestion, though heuristically useful, hardly yields a usable criterion.
- 7 Thus it is a mistake to conceive the regress as a *temporal* regress, as it would be if each justifying argument had to be explicitly given before the belief in question was justified.
 - 8 Obviously these views could be combined, with different instances of the regress being handled in different ways. I will not consider such combined views here. In general, they would simply inherit all of the objections pertaining to the simpler views.
 - 9 Peirce seems to suggest a virtuous regress view in "Questions concerning Certain Faculties Claimed for Man," *Collected Papers V*, pp. 135–55. But the view is presented metaphorically and it is hard to be sure exactly what it comes to or to what extent it bears on the present issue.
 - 10 The original statement of the non-linear view was by Bernard Bosanquet in *Implication and Linear Inference* (London, 1920). For more recent discussions, see Harman, *Thought*; and Nicholas Rescher, "Foundationalism, Coherentism, and the Idea of Cognitive Systematization," *The Journal of Philosophy* 71 (1974), pp. 695–708.
 - 11 I have attempted to show how a coherence view might be defended against the most standard of these objections in "The Coherence Theory of Empirical Knowledge," *Philosophical Studies* 30 (1976), pp. 281–312.
 - 12 The presumption against a skeptical outcome is strong, but I think it is a mistake to treat it as absolute. If no non-skeptical theory can be found which is at least reasonably plausible in its own right, skepticism might become the only rational alternative.
 - 13 For some useful distinctions among these terms, see William Alston, "Varieties of Privileged Access," *American Philosophical Quarterly* 8 (1971), pp. 223–41.
 - 14 For discussions of weak foundationalism, see Bertrand Russell, *Human Knowledge* (New York: Simon and Schuster, 1949), part II, ch. II, and part V, chs. 6 and 7; Nelson Goodman, "Sense and Certainty," *Philosophical Review* 61 (1952), pp. 160–7; Israel Scheffler, *Science and Subjectivity* (New York, 1967), chapter V; and Roderick Firth, "Coherence, Certainty, and Epistemic Priority," *The Journal of Philosophy* 61 (1964), pp. 545–57.
 - 15 How good a reason must one have? Presumably some justification accrues from any reason which makes the belief even minimally more likely to be true than not, but considerably more than this would be required to make the justification adequate for knowledge. (See note 3, above.) (The James-Clifford controversy concerning the "will to believe" is also relevant here. I am agreeing with Clifford to the extent of saying that epistemic justification requires some positive reason in favor of the belief and not just the absence of any reason against.)
 - 16 For a similar use of the notion of epistemic irresponsibility, see Ernest Sosa, "How Do You Know?" *American Philosophical Quarterly* II (1974), p. 117.
 - 17 In fact, the premises would probably have to be true as well, in order to avoid Gettier-type counterexamples. But I shall ignore this refinement here.
 - 18 On a Carnap-style *a priori* theory of probability it could, of course, be the case that very general empirical propositions were more likely to be true than not, i.e. that the possible state-descriptions in which they are true outnumber those in which they are false. But clearly this would not make them likely to be true in a sense which would allow the detached assertion of the proposition in question (on pain of contradiction), and this fact seems to preclude such justification from being adequate for knowledge.
 - 19 Chs 11–13. Bracketed page references in this section are to this book.
 - 20 Fred I. Dretske, *Seeing and Knowing* (London: Routledge and Kegan Paul, 1969), ch. III, especially pp. 126–39. It is difficult to be quite sure of Dretske's view, however, since he is not concerned in this book to offer a general account of knowledge. Views which are in some ways similar to those of Armstrong and Dretske have been offered by Goldman and by Unger. See Alvin Goldman, "A Causal Theory of Knowing," *The Journal of Philosophy* 64 (1967), pp. 357–72; and Peter Unger, "An Analysis of Factual Knowledge," *The Journal of Philosophy* 65 (1968), pp. 157–70. But both Goldman and Unger are explicitly concerned with the Gettier problem and not at all with the regress problem, so it is hard to be sure how their views relate to the sort of externalist view which is at issue here.
 - 21 On the one hand, Armstrong seems to argue that it is *not* a requirement for knowledge that the believer have "sufficient evidence" for his belief, which sounds like a rejection of the adequate-justification condition. On the other hand, he seems to want to say that the presence of the external relation makes it rational for a person to accept a belief, and he seems (though this is not clear) to have *epistemic* rationality in mind; and there appears to be no substantial difference between saying that a belief is epistemically rational and saying that it is epistemically justified.
 - 22 One way to put this point is to say that whether a belief is likely to be true or whether in contrast it is an accident that it is true depends significantly on how the belief is described. Thus it might be true of one and the same belief that it is "a belief connected in a law-like way with the state of affairs which it describes" and also that it is "a belief adopted on the basis of no apparent evidence"; and it might be

- likely to be true on the first description and unlikely to be true on the second. The claim here is that it is the believer's own conception which should be considered in deciding whether the belief is justified. (Something analogous seems to be true in ethics: the moral worth of a person's action is correctly to be judged only in terms of that person's subjective conception of what he is doing and not in light of what happens, willy-nilly, to result from it.)
- 23 Notice, however, that if beliefs standing in the proper external relation should happen to possess some subjectively distinctive feature (such as being spontaneous and highly compelling to the believer), and if the believer were to notice empirically, that beliefs having this feature were true a high proportion of the time, he would then be in a position to construct a justification for a new belief of that sort along the lines sketched at the end of section II. But of course a belief justified in that way would no longer be basic.
- 24 I suspect that something like the argument to be given here is lurking somewhere in Sellars' "Empiricism and the Philosophy of Mind" (reprinted in Sellars, *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1963), pp. 127–96), but it is difficult to be sure. A more recent argument by Sellars which is considerably closer on the surface to the argument offered here is contained in "The Structure of Knowledge," his Machette Foundation Lectures given at the University of Texas in 1971, in Hector-Nerl Casteneda (ed.), *Action, Knowledge, and Reality: Critical Studies in Honor of Wilfrid Sellars* (Indianapolis, 1975), Lecture III, sections III–IV. A similar line of argument was also offered by Neurath and Hempel. See Otto Neurath, "Protocol Sentences," tr. in A. J. Ayer (ed.), *Logical Positivism* (New York, 1959), pp. 199–208; and Carl G. Hempel, "On the Logical Positivists' Theory of Truth," *Analysis*, 2 (1934–5), pp. 49–59. The Hempel paper is in part a reply to a foundationist critique of Neurath by Schlick in "The Foundation of Knowledge," also translated in Ayer, *Logical Positivism*, pp. 209–27. Schlick replied to Hempel in "Facts and Propositions," and Hempel responded in "Some Remarks on 'Facts' and Propositions," both in *Analysis* 2 (1934–5), pp. 65–70 and 93–6, respectively. Though the Neurath–Hempel argument conflates issues having to do with truth and issues having to do with justification in a confused and confusing way, it does bring out the basic objection to givenism.
- 25 Chisholm, "Theory of Knowledge," in Chisholm et al., *Philosophy* (Englewood Cliffs, NJ: Prentice-Hall, 1964), pp. 270ff.
- 26 Bracketed page references in this section will be to this book.
- 27 Quinton does offer one small bit of clarification here, by appealing to the notion of ostensive definition and claiming in effect that the sort of awareness involved in the intuitive justification of a basic belief is the same as that involved in a situation of ostensive definition. But such a comparison is of little help, for at least two reasons. First, as Wittgenstein, Sellars, and others have argued, the notion of ostensive definition is itself seriously problematic. Indeed, an objection quite analogous to the present one against the notion of a basic belief could be raised against the notion of an ostensive definition; and this objection, if answerable at all, could only be answered by construing the awareness involved in ostension in such a way as to be of no help to the foundationist in the present discussion. Second, more straightforwardly, even if the notion of ostensive definition were entirely unobjectionable, there is no need for the sort of awareness involved to be *justified*. If all that is at issue is learning the meaning of a word (or acquiring a concept), then justification is irrelevant. Thus the existence of ostensive definitions would not show how there could be basic beliefs.
- 28 Notice, however, that to reject an epistemological given does not necessarily rule out other varieties of givenness which may have importance for other philosophical issues. In particular, there may still be viable versions of givenness which pose an obstacle to materialist views in the philosophy of mind. For useful distinctions among various versions of givenness and a discussion of their relevance to the philosophy of mind, see James W. Cornman, "Materialism and Some Myths about Some Givens," *The Monist* 56 (1972), pp. 215–33.
- 29 Compare the Husserlian notion of a "pre-predicative awareness."
- 30 It is interesting to note that Quinton seems to offer an analogous critique of givenness in an earlier paper, "The Problem of Perception," reprinted in Robert J. Swartz (ed.), *Perceiving, Sensing, and Knowing* (Garden City, NY: Doubleday, 1965), pp. 497–526; cf. especially p. 503.
- 31 For a discussion of such a coherence theory, see my paper cited in note 11, above.

Reflective Knowledge in the Best Circles

Ernest Sosa

Is the existence of external things just an article of faith? Certainly not, says G. E. Moore,¹ who offers us a proof, thus aiming to remove Immanuel Kant's "scandal to philosophy."

Moore's proof

Here is a hand (a real, flesh and bone hand).

Therefore, there is at least one external thing in existence.²

According to Moore, his argument meets three conditions for being a proof: first, the premise is different from the conclusion; second, he knows the premise to be the case; and, third, the conclusion follows deductively (*ibid.*, pp. 144–5). Further conditions may be required, but he evidently thinks his proof would satisfy these as well.

I Moore's Proof

As Moore is well aware, many philosophers will feel he has not given "any satisfactory proof of the point in question" (*ibid.*, p. 147). Some, he believes, will want the premise itself proved. But he does not try to prove it, and does not believe it can be proved. Proving that here is a hand requires proving one is awake, and this cannot be done.

Does Moore adequately answer the skeptic? Many have denied it for the reason that he fails to rule out a crucial possibility: that our faculties are leading us astray – for example, that we are dreaming. Aware of this objection, Moore grants, in

"Certainty,"³ that to know he is standing, he must know he is awake. The point "cuts both ways," however, and he would prefer to conclude that he does know he is awake since he does know he is standing. This has persuaded nearly nobody. On the contrary, some have thought him committed to an argument, *M* below, like the following:

Argument *A*

A1. This map is a good guide to this desert.

A2. According to the map, an oasis lies ahead.

A3. Therefore, an oasis lies ahead.

Argument *M*

M1. My present experience is a veridical guide to reality (and I am not dreaming).

M2. My present experience is as if I have a hand before me.

M3. Therefore, here (before me) is a hand.

When challenged on premise A1, our desert dullard responds: "I must know A1, since the only way I could know A3 is through argument *A*, and I do know A3." Is this a just comparison? Is Moore's response to the skeptic relevantly similar?⁴

If Moore depends on argument *M* for his knowledge of M3, his response seems like the dullard's. The dullard is wrong to respond as he does. He must say how he knows his premise without presupposing that he already knows the conclusion. And Moore would seem comparably wrong in the analogous response to the skeptic. In explaining how he knows M1, he must not presuppose that he already knows M3.

Does Moore depend on argument *M* for his knowledge of M3? There is reason to think that

Originally published in *The Journal of Philosophy* XCIV, 8 (1997), pp. 410–30.

he does not, given his emphatic acknowledgment that he cannot *prove* M3. After all, *M* would seem a proof of M3 just as good as Moore's own "proof of an external world." Moore concedes, in effect, that, if he does not know that he is not dreaming, *then* he does not know of the hand before him. But that is not necessarily because he takes himself to know M3 only through *M* or any other such argument. In any case, even if he is relying on some such argument, which would require making that concession, the defender of common sense has other options.

One might, after all, make that concession only because of the following "principle of exclusion" (PE):

If one is to know that *h*, then one must exclude (rule out) every possibility that one knows to be incompatible with one's knowing that *h*.

As Moore grants explicitly, the possibility that he might be just dreaming is incompatible with his knowing (perceptually) that he has a hand before him. And this, in combination with PE, is quite sufficient to explain his concession above.

Suppose Moore is not depending on argument *M* for his knowledge of M3. Although he recognizes his need to know he is not dreaming, suppose that is only because he accepts PE, our principle of exclusion. Then the sort of ridicule cast on the dullard is misdirected against Moore. What is more, it is not even clear that Moore must know *how* he knows he is not dreaming if he is to know M3. That is not entailed by the application of the principle of exclusion. All that follows from such application is that Moore must know *that* he is not dreaming, not that he must know *how* he knows this.

In fact, however, the historical Moore did rely on something very much like argument *M* (more on this below). So is he not, after all, exposed to the damaging comparison with the desert dullard? Not at all. There seems to be no good reason why, in responding to the skeptic, Moore must *show how* he knows he is not dreaming. Of course, his response to the skeptic would be enhanced if he could show that. But it now seems not properly subject to ridicule even if he is not then in a position to show how he knows he is not dreaming. The question he is addressing is *whether* he knows that he is not dreaming, and, at most, by extension, what grounds he might have for his answer to that question, in answering which he does not, nor

need he, also answer the question of *how* he knows himself to be awake and not dreaming.

It might be replied that one cannot know that here is a hand if one's belief rests on the unproved assumption that one is awake. According to Moore, however, things that cannot be proved might still be known. Besides, even though he cannot prove that he is awake, he has "conclusive evidence" for it. Unfortunately, he cannot state his evidence, and the matter is left in this unsatisfactory state at the end of "Proof of an External World." But Moore has more to say in another paper of the period, "Four Forms of Skepticism."⁵ There, he takes himself to know for sure about the hand before him, and takes this knowledge to be based on an inductive or analogical argument. We are told that introspective knowledge of one's own sensory experience, unlike perceptual knowledge of one's physical surroundings, can be immediate. While agreeing with Bertrand Russell that one cannot know *immediately* that one sees a hand, Moore thinks, contra Russell, that he can know it for certain. And he disagrees with Russell more specifically in allowing knowledge for certain about his hand through analogical or inductive reasoning from premises known introspectively.

It is doubtful, however, that any allowable form of inference – whether deductive, inductive, or analogical – will take us from the character of our experience to the sort of knowledge of our surroundings we ordinarily claim.

Familiar skeptical scenarios – dreaming, evil demon, brain in a vat, and the like – show that our experience prompts but does not logically entail its corresponding perceptual beliefs. Experience as if there is a fire before us does not entail that there is a fire there, experience as if here is a hand does not entail that here is a hand, and so on. Perhaps what is required for one's beliefs and experiences to have certain contents entails that these could not possibly be *entirely* false or misleading. Indeed, some such conclusion follows from certain externalist and epistemic requirements on one's justified attribution of familiar contents to one's own experiences or beliefs. But even if that much is right – which is still controversial – one's experience or belief that here is a hand, or yonder a fire, might still be wildly off the mark. We cannot deduce much of our supposed knowledge of the external world from unaided premises about our experience.

As for inductive or analogical reasoning, only abductive reasoning – inference to the best

explanation – offers much promise, but it seems questionable as a solution to our problem.⁶ Suppose (a) that we restrict ourselves to data just about the qualitative character of our own sensory experience, and (b) that we view belief in a commonsensical external world as a theory postulated to explain the course of our experience. What exactly is the proposal? Is it proposed that when ordinarily we accept the presence of a hand before us, we do know, and know on the basis of an abductive inference? Or is it proposed, rather, that in such circumstances we have resources that would enable us to know if only we used those resources to make effective abductive arguments? The second, more modest, proposal is too modest, since it leaves our ordinary perceptual beliefs in a position like that of a theorem accepted through a guess or a blunder, one which we do have the resources to prove after much hard thought, but one which we have not come close to proving at the time when we are just guessing or blundering.

Even the modest proposal, moreover, seems unlikely to succeed. Could we form a rich enough set of beliefs purely about the qualitative character of our sensory experience, one rich enough to permit abductive inferences yielding our common-sense view of external reality? This seems doubtful when we consider (a) that such pure data beliefs could not already presuppose the external reality to be inferred, and (b) that the postulated common-sense “theory” of external reality must presumably meet constraints on abductive inference: for example, that the postulated theory be empirically testable and also simpler and less ad hoc than alternatives (for example, George Berkeley’s). These requirements plausibly imply that our data must go beyond detached observations, and include some acceptable correlations. Yet these correlations are unavailable if we restrict ourselves to beliefs about the character of our experience.⁷ Most especially are they unavailable, and most especially is the postulated inference implausible, when our database is restricted, as it is by Moore, to introspectively known facts of one’s own then-present subjective experience and to directly recalled facts of one’s own earlier experience. (If deprived of the epistemic resources of testimony and of retentive memory – except insofar as such resources can be validated by reason-cum-introspection, which is not very far if at all – then there is precious little we can any longer see ourselves as knowing, thus deprived.)

Accordingly, the skeptic has a powerful case against Moore’s claim that our knowledge of the external world is based on an inductive or analogical inference from such information about our experience. It is not realistic to suppose that we consciously make such inferences in everyday life. It is less implausible to conceive of such inferences as implicit and/or dispositional, but even this strains belief. Besides, even granted that we make such inferences if only implicitly, do they yield simpler and less ad hoc hypotheses than alternatives? That is far from clear; nor do such hypotheses seem empirically testable and credible simply as explanations of the purely qualitative character of our then-present or directly recalled experience.

Having reached a dead end, let us have some second thoughts on Moore’s view of perceptual beliefs as inferential, whereby he joined a venerable tradition, along with Russell himself. If perceptual knowledge is thus mediate and inferential, what knowledge can qualify as immediate and foundational? Modern philosophy begins with René Descartes’s canonical answer to this question.⁸

II Descartes’s Circles

Descartes had two circles, not only the big famous one involving God as guarantor of our faculties, but also a smaller one found in the second paragraph of his third meditation, where he reasons like this:

I am certain that I am a thinking being. Do I not therefore also know what is required for my being certain about anything? In this first item of knowledge there is simply a clear and distinct perception of what I am asserting; this would not be enough to make me certain of the truth of the matter if it could ever turn out that something which I perceived with such clarity and distinctness was false. So I now seem to be able to lay it down as a general rule that whatever I perceive very clearly and distinctly is true.⁹

Yet when he looks away from particular clear and distinct items, such as the proposition that he thinks, Descartes grants that a powerful enough being could deceive him even about what seems most manifest. Descartes grants that he could be astray in his beliefs as to what he perceives or remembers, and even in taking himself to intuit

something as quite clear and distinct. This doubt must be blocked if one is to attain certainty by intuiting something as clear and distinct. Accordingly, Descartes launches the theological reflections that lead eventually to his nondeceiving God.

Even without the further boost of certainty provided by the proof of a nondeceiving God, however, Descartes takes himself to have attained some positive justification. Early in the third meditation, he takes himself to perceive clearly and distinctly that he thinks, which he takes to be what gives him the certainty that he thinks. He reasons that this clear and distinct perception would not give him such certainty if it were less than perfectly reliable, and apparently concludes from this that his clear and distinct perception is perfectly reliable. One could demand how he knows all these things: How can he be sure that he does clearly and distinctly perceive that he thinks, for one thing? How can he be sure that there is nothing else in his situation that could provide the degree of certainty involved? How can he be sure that the clarity and distinctness of his perception could not possibly provide that degree of certainty unless it were infallible? What could he say in response? Descartes might well have a uniform response to all such questions: in each case, he might appeal once again to clear and distinct perception, each of the things in question being something we are assured of by our clearly and distinctly perceiving it.

About the cogito, I wish to highlight not Descartes's answers to such questions, however, but the inference that he draws: "so I now seem to be able to lay it down as a general rule that whatever I perceive very clearly and distinctly is true." Just what is Descartes's argument in support of this general rule? Would his reasoning take the following form?

- (1) Datum: I know with a high degree of certainty that I think.
- (2) I clearly and distinctly perceive that I think, and that is the only, or anyhow the best account of the source of my knowledge that I think.
- (3) So my clear and distinct perception that I think is what explains why or how it is that I know I think.
- (4) But my clear and distinct perception could not serve as a source of that knowledge if it were not an infallibly reliable faculty.
- (5) So, finally, my clear and distinct perception must be an infallibly reliable faculty.

The move from 1 and 2 to 3 is an inference to an explanatory account that one might accept for the coherence it gives to one's view of things in the domain involved. Elsewhere, Descartes does appeal to coherence at important junctures.¹⁰ So he may be doing so here as well, although questions do arise about how Descartes views coherence. Does he accept the power of coherence to add justified certainty, and, in particular, would he claim infallibility for (sufficiently comprehensive and binding) coherence as he does for clear and distinct intuition?¹¹ In any case, the comprehensive coherence of his world view would be enhanced by an explanation of how clear and distinct perception comes to be so highly reliable, even infallible. And this is just what Descartes attempts, through his theological and other reasoning. Descartes can see that reason might take him to a position that is sufficiently comprehensive and interlocking – and thereby defensible against any foreseeable attack, no holds barred, against any specific doubt actually pressed or in the offing, no matter how slight. Unaided reason might take him to that position. Need he go any further? What is more: Might one reach a similar position while dispensing with the trappings of Cartesian rationalism?

III Circular Externalism

Compare now how Moore might have proceeded:

- (1) Datum: I know with a high degree of certainty that here is a hand.
- (2) I can see and feel that here is a hand, and that is the only, or anyhow the best account of the source of my knowledge that here is a hand.
- (3) So my perception that here is a hand is what explains why or how it is that I know (with certainty) that here is a hand.
- (4) But my perception could not serve as a source of that degree of justified certainty if it were not a reliable faculty.¹²
- (5) So, finally, my perception must be a reliable faculty.

Moore could, of course, go on to say more about the nature of the perception that assures him about the hand. He might still say that such perception involves an implicit inference from what is known immediately and introspectively, perhaps an inductive or analogical inference of some sort. That might make his view more comprehensively

coherent, but we have already seen reasons why postulating such an inference is questionable. So we focus rather on a second alternative: Moore might well take perceiving to involve no inference at all, not even implicit inference, but only transfer of light, nerve impulses, and so on, in such a way that the character of one's surroundings has a distinctive impact on oneself and occasions corresponding and reliable beliefs. This might also amount eventually to a comprehensively coherent view of one's knowledge of the external world. Its epistemologically significant features would not distinguish it in any fundamental respect from the procedure followed by Descartes.

The theme of accidentally true belief has loomed large in the epistemology of recent decades. The Gettier problem, for example, is posed by a justified belief true for reasons far removed from whatever causes it to be held and justified. Externalist conceptions of propositional knowledge focus on this theme, as do one offered by Peter Unger (nonaccidentally true belief) and one offered by Alvin Goldman (belief caused by the truth of its content). Robert Nozick's tracking account is also a conception of this sort: S knows that *p* if and only if S believes correctly that *p*, and also (in the circumstances): *both* it would have been true that *p* only if S had believed it, *and* if it had not been true that *p*, then S would not have believed it.¹³

Why are these conceptions of knowledge of special interest to us here? Because each offers a way to explain how one can know that *p* without reasoning from prior knowledge. The key idea exploited is this: you can know something noninferentially so long as it is no accident or coincidence that you are right.

Both the tracking and the causal accounts defensibly require a special nonaccidental connection between the belief and the fact believed. Nevertheless, in each case other levels of accidentality remain. Suppose I fancy myself a connoisseur of tomato ripeness, but suffer from a rare form of color blindness that precludes my discerning nearly any shade of red except that displayed by this particular tomato. Therefore, my judgments of tomato ripeness are in general apt to be right with no better than even chance. But when it is the particular (and rare) shade of red now displayed, then I am nearly infallible. Oblivious to my affliction, however, I issue judgments of tomato ripeness with abandon over a wide spectrum of shades of red. Assuming that, unknown to me, the variety

of tomato involved always ripens with this shade of red, then my belief that this tomato is ripe is in step with the truth, and arguably satisfies the requirements of Unger, Goldman, and Nozick. But, again, it is nevertheless in some relevant sense or respect only an accident that I am right in my belief.¹⁴ We need a clearer and more comprehensive view of the respects in which one's belief must be nonaccidentally true if it is to constitute knowledge.

Unaided, the tracking or causal requirements proposed suffer from a sort of tunnel vision. They permit too narrow a focus on the particular target belief and its causal or counterfactual relation to the truth of its content. Just widening our focus will not do, however, if we widen it only far enough to include the process that yields the belief involved. We need an even broader view.

IV Virtue Epistemology

[When] . . . thought is concerned with study, not with action or production, its good or bad state consists [simply] in being true or false. For truth is the function of whatever thinks.

Hence the function of each of the understanding parts is truth; and so the virtue of each part will be the state that makes that part grasp the truth most of all.¹⁵

Virtue epistemology is distinguished by its emphasis on the subject as seat of justification. In order to qualify as knowledge, a belief must be "apt," epistemically so, in a strong sense that goes beyond its being just a belief that coheres well within the subject's perspective. The "tracking account" (Nozick) sees here little more than a claim about that belief's counterfactual relation to the truth of what is believed. "Reliable-indicator" accounts require rather that the belief itself or the reasons for it have properties nomically sufficient for its truth (David Armstrong, Marshall Swain¹⁶). "Reliable-process" accounts focus instead on the cognitive process, beneath the skin, that yields the belief, and on the truth ratio in the products of that process, actual and counterfactual (Goldman).

It is rather the subject and her cognitive virtues or aptitudes which hold primary interest for virtue epistemology. Consider the athletic virtues of a tennis champion. When we say that a shot is not just a winning shot but a skillful one, we imply a

V Virtue and Coherence

comment on the player as shotmaker. Suppose a tyro wields a racquet on a court and, unaware of the approaching ball, issues what amounts by luck to a stylish and effective backhand stroke. Such a shot might be an unreturnable winning shot, but it would manifest no real skill.

Why are we unwilling to admire a performance as “skillful” if it manifests only a fleeting, or even an instantaneous state of the agent’s? Skills, abilities, competences, aptitudes, prowess – these come and go, true enough, but they do not flit by instantaneously. Why not? Why do we tend to define these concepts so as to require such stability? We might have defined similar concepts without requiring stability. Why do we define these concepts as we do? Why have we adopted these and not others? Should one not expect that, other things being equal, the more clearly useful a concept is to us, the more likely it is that we shall retain it? People need to know who are dependable members of their group – this is a *kind* of thing we need to monitor in a great variety of contexts, with a great variety of objectives. Cooperative success depends on the group’s ability to monitor people’s aptitudes and ineptitudes. So it is no surprise that the sorts of aptitudes (skills, competences, virtues) that we recognize and admire are those which linger stably.¹⁷

To praise a performance as skillful or an action as right, or a judgment as wise or apt, accordingly, is to assess not only the action or the judgment, but also the reflected aptitude or character or intelligence. This is a distinctive view with versions both in epistemology and in ethics. It is distinctive in that the rightness of an action (or a choice) and the aptness (or positive epistemic status) of a belief would involve not just whether the performance is *optimific* (if an action) or *true* (if a belief); nor just whether a good enough procedure was followed, perhaps accidentally, in arriving at that choice, or whether a good enough cognitive process chanced to lead to the belief; nor even just whether a rule, in effect, somehow demands that choice or that belief in those circumstances. Our virtue epistemology and virtue ethics focus rather on the agent and cognizer. When the agent’s actions are said to be *right* and the cognizer’s beliefs *knowledge*, we speak implicitly of the virtues, practical or intellectual, seated in that subject, which (a) give rise to that action or belief, adding to the subject’s worth as agent or cognizer, and (b) make him reliable and trustworthy over an interesting spread of possible choices or beliefs, and circumstances.

Can we explain what distinguishes a system of beliefs (and experiences) that, internally regarded, is intellectually virtuous and admirable? Presumably, our explanation would involve the system’s explanatory coherence, overall simplicity, lack of ad hoc epicycles, and so on. Philosophical mythology contains creatures who excel in all such respects, however, though their beliefs fall short of being knowledge even when true. Take the brain in a vat, for example, or the victim of the evil demon. An adult recently envatted, or victimized by the demon, can be indistinguishable from the best of us in respect of the comprehensiveness and coherence of their beliefs and experiences. Even when right about enviring objects and events, such a victim’s beliefs are far from being knowledge.

Might such comprehensive internal coherence exhaust all cognitive or intellectual virtue, at least when it comprises not only beliefs but also experiences? To assent here would be overbold even for a rationalist. Yet coherence is, of course, valued not only by philosophers but by the reflective more generally. One also wants faculties and virtues beyond reflective, coherence-seeking reason: perception, for example, and memory. Equally, internal coherence goes beyond such faculties, and requires reason, which counts for a lot in its own right. But why should that be so, if comprehensive coherence is no guarantee of truth, if the internal coherence enjoyed by the envatted yields little if any truth?

Compare first reason with memory. Input beliefs are required for retentive memory and inferential reason, which then yield beliefs as outputs. Retentive memory yields again the input belief itself, while inferential reason yields a new belief. Even the most excellent transmission faculty will not guarantee the truth of its output, which will depend not only on the transmission but also on the inputs. But our transmission faculties are valuable even so, if only because they combine with other faculties to increase vastly the total yield of true beliefs.

How does internal coherence, of little significant epistemic value in itself, become more valuable when combined with external aptness? Coherence-seeking inferential reason, like retentive memory, is of epistemic value when combined with externally apt faculties of perception, because when so combined it, like retentive memory, gives

us a more comprehensive grasp of the truth than we would have in its absence.

Good perception is in part constituted by certain transitions from experiences to corresponding beliefs – as is the transition from the visual experience characteristic of a tomato seen in good light to belief in the tomato. Other such transitions help constitute good introspection, as when one's headache prompts awareness of it as a headache. Finally, if the comprehensive coherence of one's system of beliefs is at least in part responsible for its constitution and persistence, it thereby manifests a virtuous faculty of reason. Such comprehensive coherence is not just mechanical, but must reflect appropriate sensitivity to factors like ad hocness, simplicity, and explanatory power. And it must include, not only belief/belief connections, but also experience/belief connections constitutive of good perception, and conscious-state/belief connections constitutive of good introspection. This broader conception of the coherence of one's mind involves not only the logical, probabilistic, and explanatory relations among one's first-order beliefs, but also coherence between these beliefs and one's sensory and other experiences, as well as comprehensive coherence between first-order experiences, beliefs, and other mental states, on one side, and beliefs about first-order states, on the other.

We may well ask about certain aspects of broad coherence – for example, the experience/belief transitions as well as the enumerative and abductive inferences involved – why these should be viewed as adding to the subject's intellectual worth or merit. "Because it is truth conducive," or at least in good measure for that reason, we are told, "because it increases the likelihood that the subject will have true beliefs and avoid false ones." But that is obviously false of victims in skeptical scenarios, who nevertheless are internally coherent, and even epistemically justified.

Although that seems undeniable, we can perhaps understand it comfortably if we distinguish two sorts of epistemic justification: (a) *S* is "same-world justified" in believing *P* in world *W* if and only if *S* believes *P* in *W* in virtue of a faculty that in *W* is truth conducive; and (b) *S* is "actual-world justified" in believing *P* in world *W* if and only if *S* believes *P* in *W* in virtue of a faculty that in our actual world is truth conducive.

Such relativizing and contextualizing is familiar enough in ordinary thought and speech. Here, it enables us to combine the following theses: (a) our

broad coherence is necessary for the kind of reflective knowledge traditionally desired; and (b) such broadly coherent knowledge is desirable because in our actual world it helps us approach the truth and avoid error. This is not to deny that there is a kind of "animal knowledge" untouched by broad coherence. It is rather only to affirm that beyond "animal knowledge" there is a better knowledge. This reflective knowledge does require broad coherence, including one's ability to place one's first-level knowledge in epistemic perspective. But why aspire to any such thing? What is so desirable, epistemically, about broad coherence? Broad coherence is desirable because it yields integrated understanding, and also because it is truth conducive, even if in a demon world broad coherence fails this test and is not truth conducive. Even so, we can still regard broad coherence as intellectually valuable and admirable so long as we do not regard our world as such a world.

We are now in just the position of arch-internalist Descartes. Consider the following passage:

The fact that an atheist can be "clearly aware that the three angles of a triangle are equal to two right angles" is something I do not dispute. But I maintain that this awareness of his [*cognitionem*] is not true knowledge [*scientia*], since no act of awareness that can be rendered doubtful seems fit to be called knowledge [*scientia*]. Now since we are supposing that this individual is an atheist, he cannot be certain that he is not being deceived on matters which seem to him to be very evident (as I fully explained). And although this doubt may not occur to him, it can still crop up if someone else raises the point or if he looks into the matter himself. So he will never be free of this doubt until he acknowledges that God exists.¹⁸

Descartes considers reasons to doubt, not only one's faculties of perception, memory, and introspection, but even one's faculty of intuitive reason, by which one might know that $3 + 2 = 5$, that if one thinks one exists, and the like. He defends against such doubts by coherence-inducing theological reasoning that yields an epistemic perspective on himself and his world, in terms of which he can feel confident about the reliability of his faculties, including the very faculties employed in arriving, via a priori theological reasoning, at that perspective on himself and his world, the perspect-

ive that enables him to see his world as epistemically propitious.¹⁹

In structure, virtue perspectivism is thus Cartesian, though in content it is not. Radical rationalism admits only (rational) intuition and deduction (along with memory) as its faculties of choice (or anyhow of top choice) and wishes to validate all knowledge in terms of these faculties; thus the Cartesian grand project. Virtue perspectivism admits also perception and introspection, along with intuition and deduction, as well as inductive and abductive reasoning. Gladly using all such faculties, it also accepts through testimony the aid of one's epistemic community. Fortunately, the overview thus attained inspires confidence in the means used.

Rejected as viciously circular by Descartes's critics, and by many today, our procedure does present a troubling aspect of circularity. A closer look, however, may show this to be only an illusion.²⁰

VI Epistemic Circularity: What is the Problem?

"I think, therefore I am," says Descartes, adding: "Here at last is something I really know. But what is it about this knowledge that makes it knowledge? As far as I can see, it is knowledge because it is a clear and distinct intuition. But it would not be real knowledge unless such intuition were reliable. So I can already lay it down as a general rule that clear and distinct intuition *is* reliable."

"Here is a hand," says Moore, adding: "Here is something I really know. But what gives me this knowledge? As far as I can see, it is knowledge in virtue of being a deliverance of perceptual experience. But it would not be knowledge if I were dreaming. So I can already conclude that I am *not* dreaming."

Descartes goes on to buttress the reliability of his rational intuition by developing a theology through vigorous use of that very rational intuition. Moore can similarly appeal to what he knows about his reliable senses on the basis largely of those very senses.

But is not any such reasoning circular? Yes, circular it does seem to be, "*epistemically* circular," let us say. But is it *viciously* circular? Skeptics through the ages have attacked it as such. Sextus Empiricus already uses the tropes of Agrippa in order to develop the so-called *diallelus*, or "prob-

lem of the criterion." Many have followed his lead in a long tradition. Today, skepticism-cum-relativism has spread beyond epistemology and ethics, beyond philosophy, and even beyond the academy, and its champions often wield circularity as a weapon. But, again: Is such circularity vicious? To say that it is vicious, in the present context, is to say that it is somehow bad, intellectually bad, that it puts us in a situation that is somehow intellectually unsatisfactory. When we ask how the circularity is vicious, therefore, what we want to know is just how it puts us in an unsatisfactory state: When we reason in the way alleged to be viciously circular, wherein lies the defect in our reasoning or in the resulting state?

Largely through the use of rational intuition, Descartes supports the view that rational intuition is reliable and that through its exercise he knows that he thinks and exists. Largely through the use of perception, Moore could support the view that perception is reliable, that he is not misled by a dream, and that through the exercise of perception he knows of the hand before him. If a crystal ball claims itself to be reliable, then, largely through the crystal ball, a crystal-ball gazer could support the view that such gazing is reliable, that it is rarely misleading, and that through the crystal ball he can foretell the future.

Epistemic circularity is vicious, it might be said, because it would make the gazer as well justified as Descartes or Moore. *Since* there is no way to support adequately the view that intuition is reliable, or that perception is reliable, without employing those very faculties; and *since* the same goes for memory, deduction, abduction, and testimony; *therefore*, there is no way to arrive at an acceptable theory of our knowledge and its general sources.

Perhaps that shows only how defective is the attempt to develop such a general theory of one's knowledge and its sources. There is an easy way to avoid the intellectual discomfort of having to use a faculty in answering the question whether that faculty is reliable, namely, not to ask the question. Call this the *avoidance strategy*.

Of course, we shall hardly lack company if we avoid philosophy because we find it frustrating. But the avoidance strategy that I wish to consider is not just a rejection of what seems too difficult for one's own intelligence. The implication of the avoidance strategy is not that there is something lacking in one's intelligence but that there is something wrong with the questions avoided.

Much might indeed be wrong with our very general, philosophical questions. Many find them too abstract, too impractical, too useless, and so on. But these are not the concerns of my avoidance strategist. He is, after all, a philosopher. His concern is not that the questions are just too hard for his intelligence, nor is it their abstractness, impracticality, or uselessness. He would hardly have gone into philosophy with such concerns, nor are they his concerns now. Difficulty, abstractness, impracticality, and uselessness are not in his view disqualifying drawbacks.

Why then should one as philosopher avoid questions of epistemology, such as those about the reliability of one's faculties? These questions become pressing with the realization that only if they reliably yield truth can our faculties yield knowledge. This is not just a commitment peculiar to contemporary reliabilism. Indeed, it is found already in Descartes, who, as we have seen, also stresses that intuition (and clear and distinct perception) yields knowledge only if reliable.

Consider again our principle of exclusion:

If one is to know that p , then one must exclude (rule out) every possibility that one knows to be incompatible with one's knowing that p .²¹

(By "excluding" here I mean "knowing not to be the case.")

On the basis of PE, we can see that in order to know that p , one must know that the faculties employed in arriving at one's belief that p are reliable faculties. After all, just consider the possibility that one's operative faculties were unreliable. That is surely a possibility generally known to be incompatible with attaining knowledge through them. Unreliable mechanisms of belief acquisition will not yield knowledge.

If the principle of exclusion is right, therefore, one cannot possibly know that p unless one knows that the faculties involved are reliable. But this is just the sort of knowledge that we seem able to attain only through epistemically circular reasoning.

One might, of course, question the principle of exclusion.²² One might hold that in order to know that p , one's pertinent faculties need only *be* reliable; one need not *know* them to be reliable.

One might, for example, appeal to a conception of knowledge as mere tracking. One might grant Richard Rorty that causation should not be confused with justification, while joining Nozick in

taking tracking as the essence of knowledge. "To know is just to mirror (or to track) nature. Justification is quite another matter. Justification of some sort may well require the principle of exclusion. Thus, it may be that in order to be justified in believing that p , one must exclude every possibility one knows to be incompatible with one's knowing that p . But such justification is *not* required for simple knowledge."

That response seems essentially right. What is more, even Descartes would agree. For Descartes, you will recall, our knowledge that our faculties are reliable, even our faculty of reason, depends on our knowledge of God's epistemic good will. Yet, as we have seen, Descartes grants explicitly that the atheist mathematician can know some mathematics.

The knowledge of an atheist is said to be *cognitio*, however, a second-class accomplishment by comparison with *scientia*. *Scientia*, by contrast, does require relevant knowledge of one's reliability. Only thus can one repel doubts about the possible unreliability of one's faculties. Only thus can one exclude a possibility evidently incompatible with one's knowing that p , namely, the possibility that only unreliable faculties yield one's belief.

By analogy, we can more generally distinguish *animal* knowledge, which requires only that one track reality, on one hand, and *reflective* knowledge, on the other, which in addition requires awareness of how one knows, in a way that precludes the unreliability of one's faculties. Unlike Descartes's *cognitio* and *scientia*, our more general animal and reflective knowledge do not require infallible reliability, but only a high level of reliability.²³ The avoidance strategy now has not only the cost of suppressing philosophical curiosity about knowledge. We can now see how it also precludes first-level reflective knowledge, and of course *scientia*.

Given these costs, what again counts in favor of avoidance? So far we have been told that we must avoid epistemic circularity because it entails arriving at a generally positive view of one's faculties only by use of those very faculties. But why should that be frustrating when it is the inevitable consequence of its generality? So far the answer is only that the superstitious crystal gazer could reason analogously and with equal justification in defense of his own perspective. How damaging is this?

Suppose we grant the gazer epistemic justification and internal coherence equal to our own. Still,

internal coherence is clearly insufficient. Is that not obvious in view of paranoia, hypochondria, and similar psychoses? Logical brilliance permits logical coherence but does not even ensure sanity, much less general epistemic aptitude. There are faculties other than reason whose apt functioning is also crucial to the subject's epistemic welfare.

In light of that result, why not distinguish between the gazers and the perceivers in that, though both reason properly and attain thereby coherence and justification, only the perceivers' beliefs are epistemically apt and constitute knowledge?

On this view, the crystal gazers differ from the perceivers in that gazing is not reliable while perceiving is. So the theory of knowledge of the perceivers is right, that of the gazers wrong. Moreover, the perceivers can know their theory to be right when they know it in large part through perception, since their theory is right and perception can thus serve as a source of knowledge. The gazers are, by hypothesis, in a very different position. Gazing, being unreliable, cannot serve as a source of knowledge. So the perceivers have a good source or basis for their knowledge, but the gazers, lacking any such source or basis, lack knowledge.

Still one might insist that the perceivers should not be so smug. They should still feel acute discomfort and intellectual frustration. This I find a very widely shared view, in epistemology and mutatis mutandis, far beyond. According to Barry Stroud,²⁴ the perceivers can at best reach a position where they can affirm the conditional proposition that if their perception is reliable, then they know. And he has recently reemphasized what is essentially the same thesis as follows:

Sosa's "externalist" could say at most: "If the theory I hold is true, I do know or have good reason to believe that I know or have good reason to believe it, and I do understand how I know the things I do." I think . . . we can see a way in which the satisfaction the theorist seeks in understanding his knowledge still eludes him. Given that all of his knowledge of the world is in question, he will still find himself able to say only "I might understand my knowledge, I might not. Whether I do or not all depends on how things in fact are in the world I think I've got knowledge of."²⁵

It is not easy to understand this position, however. If our perceivers believe (a) that their perception,

if reliable, yields them knowledge, and (b) that their perception is reliable, then why are they restricted to affirming only the conditional, *a*, and not its antecedent, *b*? Why must they wonder whether they understand their relevant knowledge? Indeed, to the extent that they are really convinced of both *a* and *b*, it would seem that, far from being logically constrained to wondering whether they know, they are, on the contrary, logically constrained from so wondering. After all, first, if you are really certain that *p*, then you cannot well consider whether you know it without thinking that you do. Moreover, second, is it not incoherent to be convinced that *p* and yet wonder whether *p*?

In sum, I see no sufficient reason to settle either for irresolvable frustration or for the avoidance strategy. The main argument we have seen for that depends on the claim that, if we allow the circular defense offered for externalist epistemology, then the gazers turn out no less epistemically justified than the perceivers. In a sense that is true: but then in a sense they are equally internally justified, equally coherent. Nevertheless, their beliefs are *not* equally apt in all epistemically relevant respects. Perception is, of course, reliable while gazing is not. Therefore, the perceivers are right and apt both in their particular perceptual beliefs, at least generally, and in their theory of knowledge – for it all rests in large measure on their reliable perception. By contrast, the gazers are wrong and inapt both in their particular gaze-derived beliefs and in their theory of knowledge – for it all rests on their unreliable gazing. Moreover, I see no reason why the perceivers must be restricted to affirming only the conditional that, if perception is reliable, then they know. I see no reason why they cannot also affirm the antecedent, why they cannot believe, both rationally and aptly, that perception is reliable and does enable them to know.

VII Circles Beyond Belief

Why require the appeal to comprehensive enough coherence for justification, an appeal that I have attributed, tentatively, to Descartes, as part of what justifies his recourse to theology in accounting for true knowledge (*scientia*)? Why not say that what justifies is that one's beliefs be caused by the gods? If the question arises – why not add that *this* belief itself is justified because it is itself caused by

the gods? – we could, of course, proceed in this simplified way without worrying about coherence or about the source of these beliefs beyond attributing them to divine agency. But that is not the way we are built, most of us: we just do not acquire such beliefs the way we do acquire beliefs willy-nilly when we open our eyes in good light. But what if we were built that way? Would we then be justified in having such beliefs, and in explaining our justification for having them, by their origin in divine agency? Would we then be justified to the degree and in the way in which Descartes is justified or in the way in which our imagined Moore would be justified through his appeal to a more ordinary reliabilism than that of Descartes? Internally regarded, the structure of beliefs would share prominent features in all three cases. Of course, from our Moorean, common-sense position we can object both to Cartesianism and to the invocation of the gods. These views are internally coherent, but we might still reject them as wrong. And we might be able to explain what is wrong with them, from our point of view, especially if our point of view rules out their leading ideas. But they can, for

their part, return the favor. Besides, we can anyhow imagine someone brilliant but insane, who weaves a system of immense interlocking complexity, but one wholly detached from reality as we know it commonsensically. Such a madman could object to our common-sense beliefs in a way that would seem relevantly analogous to the way in which we would object to his mad beliefs.

What all of that shows, it seems to me, is nothing more than that knowledge does not live by coherence and truth alone. Knowledge requires truth and coherence, true enough, but it often requires more: for example, that one be adequately related, causally or counterfactually, to the objects of one's knowledge, which is not necessarily ensured by the mere truth-cum-coherence of one's beliefs, no matter how comprehensive the coherence. Madmen can be richly, brilliantly coherent; not just imaginary madmen, but real ones, some of them locked up in asylums. Knowledge requires not only internal justification or coherence or rationality, but also external warrant or aptness. We must be both in good internal order and in appropriate relation to the external world.

Notes

- 1 "Proof of an External World," this vol., ch. 2.
- 2 This is, of course, a simplified version of Moore's proof.
- 3 This vol., ch. 4.
- 4 Compare Barry Stroud, *The Significance of Philosophical Scepticism* (New York: Oxford, 1984), ch. 1 and 3.
- 5 This vol., ch. 3.
- 6 For Russell the "common-sense hypothesis" of independent physical objects is "simpler" than the supposition that life is but a dream (as he explains in *The Problems of Philosophy* (New York: Oxford, 1959), ch. 2). For W. V. Quine, the "hypothesis of ordinary physical objects" is "posited" or "projected" from the data provided by sensory stimulations: "Subtracting his cues from his world view, we get man's net contribution as the difference" – *Word and Object* (Cambridge: MIT, 1960), p. 5. That Quine's position is deeply problematic is shown by Stroud, *The Significance of Philosophical Scepticism*, ch. 6.
- 7 This is argued by Wilfrid Sellars in "Phenomenalism," in his *Science, Perception, and Reality* (New York: Routledge, 1963), pp. 60–105.
- 8 The shift to discussion of Descartes may seem abrupt; however, what we find about the nature of immediate knowledge in that discussion has important implications for a position that Moore failed to

- explore. Sceptics who are willing to grant Descartes his immediate knowledge through introspection or rational intuition would need to explain exactly why perception could never yield such knowledge. (And what of memory?) The discussion of Descartes to follow is meant to highlight this issue.
- 9 J. Cottingham, R. Stoothoff, and D. Murdoch (eds), *The Philosophical Writings of Descartes* (New York: Cambridge, 1975), vol. II, p. 24.
- 10 In his *Principles of Philosophy* (part IV, art. 205), for example, he notes that, if we can interpret a long stretch of otherwise undecipherable writing by supposing that it is written in "one-off natural language," where the alphabet has all been switched forward by one letter, and so on, then this is good reason for that interpretation. There, he also argues for his scientific account of reality in terms of certain principles by claiming that "it would hardly have been possible for so many items to fall into a coherent pattern if the original principles had been false" – *The Philosophical Writings of Descartes*, vol. I, p. 290.
- 11 My attribution to Descartes is tentative because of the enormous bibliography on the "Cartesian Circle." In deference to that important tradition of scholarship, I do no more than suggest that there is logical space for an interpretation of Descartes that

- is perhaps more complex than many already tried, but that seems coherent and interesting. (I am myself convinced that this is Descartes's actual position, and defend this more fully in "How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes," *Philosophical Studies* LXXXV (1997), pp. 229–49. In "Mythology of the Given," in *The History of Philosophy Quarterly* 14 (1997), pp. 275–86, I argue for the relevance of this Cartesian strategy to issues of empirical foundations that have divided philosophers since the Vienna Circle, and have pitted, for example, Wilfrid Sellars, Richard Rorty, and Donald Davidson, on one side, against C. I. Lewis, Carl Hempel, and Roderick Chisholm, on the other. My warm thanks to Lex Newman and James Van Cleve for helpful discussion of, and further references relevant to, this way of viewing Descartes.)
- 12 Here, one would reduce Descartes's requirement of infallible certainty.
 - 13 Unger, "An Analysis of Factual Knowledge," *The Journal of Philosophy*, LXV, 6 (March 21, 1968), pp. 157–70; Goldman, "A Causal Theory of Knowing," *The Journal of Philosophy*, LXIV, 12 (June 22, 1967), pp. 357–72; Nozick, *Philosophical Explanations* (Cambridge: Harvard, 1981), ch. 3 (see this vol, ch. 10).
 - 14 For an early statement of this sort of problem, urged against Nozickian tracking, see Colin McGinn's "The Concept of Knowledge," *Midwest Studies in Philosophy*, IX (1984), pp. 529–54.
 - 15 Aristotle, *Nicomachean Ethics*, T. Irwin, trans. (Indianapolis: Hackett, 1985), 1139a27–30.
 - 16 Armstrong, *Belief, Truth, and Knowledge* (New York: Cambridge, 1973); Swain, *Reasons and Knowledge* (Ithaca: Cornell, 1981).
 - 17 The social utility of concepts is invoked occasionally to defend a proposed account of knowledge in my *Knowledge in Perspective* (New York: Cambridge, 1991); see pp. 27, 275. The form of argument involved is used insightfully in E. J. Craig's *Knowledge and the State of Nature* (New York: Oxford, 1990), where it takes center stage.
 - 18 This passage is from the Second Set of Replies as it appears in *The Philosophical Writings of Descartes*, vol. II, p. 101. I must add, however, that where this translation says that an atheist can be "clearly aware," Descartes's Latin is *clare cognoscere*.
 - 19 Although unremarked by Descartes, the role of dreams in his perception skepticism is analogous to a role assignable to paradoxes and aporias in a parallel skepticism vis-à-vis rational intuition.
 - 20 See my "Philosophical Skepticism and Epistemic Circularity," *Proceedings of the Aristotelian Society*, Supplementary Volume LXIV (1994): 263–90; and compare Stroud's response, "Scepticism, 'Externalism', and the Goal of Epistemology," pp. 291–307.
 - 21 Chapter 1 of Stroud's *The Significance of Philosophical Scepticism* is an illuminating discussion of this principle and its importance for understanding philosophical skepticism. In "How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes," I suggest how to derive it from other principles with independent plausibility.
 - 22 For one thing, as it stands it leads, apparently, to a vicious regress. But that is an illusion. After all, what PE requires one to rule out is, not every possibility incompatible with one's knowing that *p*, but rather every possibility known to be thus incompatible. Since, for one thing, knowledge requires belief, the regress is hence not infinite, nor does it seem vicious.
 - 23 This distinction figures in my *Knowledge in Perspective* (New York: Cambridge, 1991), pp. 240, 282.
 - 24 "Understanding Human Knowledge in General," in M. Clay and K. Lehrer (eds), *Knowledge and Scepticism* (Boulder: Westview, 1989), pp. 31–50, here p. 47.
 - 25 "Scepticism, 'Externalism', and the Goal of Epistemology," pp. 303–4.



PART VI

Epistemology Naturalized

Introduction

W. V. Quine is well known for urging the abandonment of epistemology, as traditionally pursued, in favor of the scientific project he calls "naturalized epistemology." Traditional epistemology, he claims, consists of two projects, the doctrinal and the conceptual project. The doctrinal project aims to deduce our material object beliefs from premises about observations, the conceptual project to reduce material object concepts to sense experience concepts. Both projects are doomed to failure. Yet epistemology, or something like it, "still goes on, though in a new setting and a clarified status." Epistemology becomes "a chapter of psychology and hence of natural science." Epistemology has always been concerned with the foundations of science, and rightly so, says Quine. But the relations between science and its foundations, the *evidence* on which it is based, viz. the totality of our sensory stimulations, are more usefully investigated scientifically, causally. It is true that scientific theories explaining the relation between stimulation and theory themselves have their source ultimately in stimulation. Yet there is no circularity here, for there is no attempt, as there is in traditional epistemology, to find a form of understanding better than our best science. We may legitimately use science to investigate its foundations.

Jaegwon Kim argues that Quine is not entitled to think of naturalized epistemology as investigating the evidential foundations of science. Quine is urging us to dispense with the normative element in epistemology, yet evidence itself is a normative concept, distinct from and irreducible to the naturalistic concepts employed in science, e.g., the concepts of stimulation, causation, law, etc. Traditional

epistemology asks what makes certain states, certain experiences and beliefs, evidence for other beliefs. Empirical psychology offers no help in answering this question, or any other question about what confers positive epistemic value. Kim goes on to provide a Davidsonian argument for the claim that Quine is not even entitled to speak of naturalized epistemology as investigating the ancestry of beliefs. For to investigate this ancestry, one must identify beliefs, but the identification of beliefs is possible only on the assumption that one's subjects are rational, where rationality here is normative and includes epistemic rationality. If Quine's "successor subject" is to avoid the normative entirely, then *contra* Quine, it is a radically different subject.

Kim ends his piece by providing an argument for the possibility of traditional normative epistemology. Epistemic properties, like all normative properties, are supervenient. If something is morally right, it is so because it possesses certain non-moral, ultimately non-normative, properties. There is thus a guarantee that there are correct normative epistemological principles, though not necessarily ones that admit of simple formulation, for there is a guarantee that there are non-epistemic, and ultimately non-normative, conditions that underlie epistemic properties.

The selection from Hilary Putnam in this section connects in specific and general ways with the debate between Quine and Kim. The specific connection is found in Putnam's argument that Quine's naturalized epistemology either eliminates the normative element in rationality, as Kim thinks it does, or reduces it to what Putnam calls "a solipsism of the present moment," the view that

Introduction

the only notion of rightness there is is the notion of what I now accept. The general connection lies in Putnam's argument that rationality can be neither eliminated nor naturalistically reduced. To eliminate it is to distort the making of statements as the mere production of noise, to distort thought to mere subvocalization. To attempt to naturalize it, for example, in the reliabilist way or in Quine's, is to misconstrue the notion of truth. It is to fall into one of two traps, absolutist metaphysical realism or solipsism of the present moment. We must realize that the notions of truth and rationality are inextricably linked: truth is a kind of idealized rationality.

In the final selection in this section, Robert Audi examines the question of whether a kind of foundationalism can adequately answer the skeptic. He answers in the affirmative. A holistic, fallibilist foundationalism can serve as the basis for the following kind of dialectical showing. (The *perceptual principle* mentioned below holds that if S has a clear sensory impression of x 's being F and on the basis of that impression believes that x is F , then this belief is justified.)

- (a) I justifiably believe both the perceptual principle and that I now have a clear impression of a bespeckled white surface before me.
- (b) On this basis, I believe that there is a bespeckled white surface before me.
- (c) My belief that there is a bespeckled white surface before me is justified.

Audi's principal concern about such showings is whether they are epistemically circular (though he does not use this terminology). Epistemic circular-

ity can be given a narrow and a broad interpretation. Under the narrow interpretation, an argument is epistemically circular if and only if at least one of the premises used to support the conclusion *relies* for its support (at least in part) on the conclusion. In the broad sense, an argument is epistemically circular just in case at least one of the premises relies for its support on a claim that in turn relies for its support on that premise itself. Thus, if one's defense of (a) adverted to claims such as (d), as a part of a track-record argument in favor of the perceptual principle, one's argument (a)–(c) would exhibit broad epistemic circularity:

- (d) I justifiably believe that: (1) I have a clear sensory impression of there being a flat brown surface before me; (2) on the basis of this impression I believe that there is a flat brown surface before me; and (3) my belief that there is such a surface before me is justified.

Audi addresses these concerns by claiming that he knows (a) (as well as (b)) non-inferentially, but that appeal to justified acceptance of instances of the perceptual principle, such as our (d), provides additional support. Thus epistemic circularity is avoided, and so, too, is the requirement that all justificatory support be linear in structure.

Audi ends his piece by considering whether his brand of foundationalism is naturalistic. He maintains that, while it is neither an empirical theory nor a theory that attempts to reduce the epistemic to the non-normative, it is in keeping with the view that epistemic properties supervene on non-normative or natural properties.

Further Reading

Armstrong, D. M., *Belief, Truth and Knowledge* (Cambridge: Cambridge University Press, 1973).

Dretske, Fred, *Knowledge and the Flow of Information* (Cambridge, MA: Massachusetts Institute of Technology Press, 1981).

Foley, Richard, "Quine and Naturalized Epistemology," *Midwest Studies in Philosophy*, vol. XIX (1994), pp. 243–61.

Goldman, Alvin, *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).

—, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA: Massachusetts Institute of Technology Press, 1992).

—, "Epistemic Folkways and Scientific Epistemology," reprinted in *Naturalized Epistemology*, 2nd

edn, ed. Hilary Kornblith (Cambridge, MA: MIT Press, 1994), pp. 219–315.

Haack, Susan, *Evidence and Inquiry: Towards Reconstruction in Epistemology* (Oxford: Blackwell, 1993), ch. 6.

Harman, Gilbert, *Change in View: Principle of Reasoning* (Cambridge, MA: Massachusetts Institute of Technology Press, 1986).

Hookway, Christopher, *Quine* (Cambridge, UK: Polity Press, 1988), pt. IV.

Kitcher, Philip, "The Naturalists Return," *Philosophical Review* CI (1992), pp. 53–114.

Kornblith, Hilary, *Inductive Inference and Its Natural Ground: An Essay in Naturalized Epistemology* (Cambridge, MA: Massachusetts Institute of Technology Press, 1993).

- , ed., *Naturalizing Epistemology*, 2nd edn (Cambridge, MA: Massachusetts Institute of Technology Press, 1994).
- Plantinga, Alvin, *Warrant and Proper Function* (Oxford: Oxford University Press, 1993).
- Quine, W. V. O., *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1992), ch. 1.
- Sosa, Ernest, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991), Part II.
- Stein, Edward, *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science* (Oxford: Oxford University Press, 1996).
- Stich, Stephen, *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation* (Cambridge, MA: Massachusetts Institute of Technology Press, 1990).
- Stich, Stephen and Nisbett, Richard, "Justification and the Psychology of Human Reasoning," *Philosophy of Science* 47 (1980), pp. 188–202.

Epistemology Naturalized

W. V. Quine

Epistemology is concerned with the foundations of science. Conceived thus broadly, epistemology includes the study of the foundations of mathematics as one of its departments. Specialists at the turn of the century thought that their efforts in this particular department were achieving notable success: mathematics seemed to reduce altogether to logic. In a more recent perspective this reduction is seen to be better describable as a reduction to logic and set theory. This correction is a disappointment epistemologically, since the firmness and obviousness that we associate with logic cannot be claimed for set theory. But still the success achieved in the foundations of mathematics remains exemplary by comparative standards, and we can illuminate the rest of epistemology somewhat by drawing parallels to this department.

Studies in the foundations of mathematics divide symmetrically into two sorts, conceptual and doctrinal. The conceptual studies are concerned with meaning, the doctrinal with truth. The conceptual studies are concerned with clarifying concepts by defining them, some in terms of others. The doctrinal studies are concerned with establishing laws by proving them, some on the basis of others. Ideally the more obscure concepts would be defined in terms of the clearer ones so as to maximize clarity, and the less obvious laws would be proved from the more obvious ones so as to maximize certainty. Ideally the definitions would generate all the concepts from clear and

distinct ideas, and the proofs would generate all the theorems from self-evident truths.

The two ideals are linked. For, if you define all the concepts by use of some favored subset of them, you thereby show how to translate all theorems into these favored terms. The clearer these terms are, the likelier it is that the truths couched in them will be obviously true, or derivable from obvious truths. If in particular the concepts of mathematics were all reducible to the clear terms of logic, then all the truths of mathematics would go over into truths of logic; and surely the truths of logic are all obvious or at least potentially obvious, i.e., derivable from obvious truths by individually obvious steps.

This particular outcome is in fact denied us, however, since mathematics reduces only to set theory and not to logic proper. Such reduction still enhances clarity, but only because of the interrelations that emerge and not because the end terms of the analysis are clearer than others. As for the end truths, the axioms of set theory, these have less obviousness and certainty to recommend them than do most of the mathematical theorems that we would derive from them. Moreover, we know from Gödel's work that no consistent axiom system can cover mathematics even when we renounce self-evidence. Reduction in the foundations of mathematics remains mathematically and philosophically fascinating, but it does not do what the epistemologist would like of it: it does not reveal the ground of mathematical knowledge, it does not show how mathematical certainty is possible.

Still there remains a helpful thought, regarding epistemology generally, in that duality of structure

Republished with permission of Columbia University Press, 562 W. 113th St, New York, NY 10025. *Ontological Relativity and Other Essays*, W. V. Quine, 1969. Reproduced by permission of the publisher via Copyright Clearance Center, Inc.

which was especially conspicuous in the foundations of mathematics. I refer to the bifurcation into a theory of concepts, or meaning, and a theory of doctrine, or truth; for this applies to the epistemology of natural knowledge no less than to the foundations of mathematics. The parallel is as follows. Just as mathematics is to be reduced to logic, or logic and set theory, so natural knowledge is to be based somehow on sense experience. This means explaining the notion of body in sensory terms; here is the conceptual side. And it means justifying our knowledge of truths of nature in sensory terms; here is the doctrinal side of the bifurcation.

Hume pondered the epistemology of natural knowledge on both sides of the bifurcation, the conceptual and the doctrinal. His handling of the conceptual side of the problem, the explanation of body in sensory terms, was bold and simple: he identified bodies outright with the sense impressions. If common sense distinguishes between the material apple and our sense impressions of it on the ground that the apple is one and enduring while the impressions are many and fleeting, then, Hume held, so much the worse for common sense; the notion of its being the same apple on one occasion and another is a vulgar confusion.

Nearly a century after Hume's *Treatise*, the same view of bodies was espoused by the early American philosopher Alexander Bryan Johnson.¹ "The word iron names an associated sight and feel," Johnson wrote.

What then of the doctrinal side, the justification of our knowledge of truths about nature? Here, Hume despaired. By his identification of bodies with impressions he did succeed in construing some singular statements about bodies as indubitable truths, yes; as truths about impressions, directly known. But general statements, also singular statements about the future, gained no increment of certainty by being construed as about impressions.

On the doctrinal side, I do not see that we are further along today than where Hume left us. The Humean predicament is the human predicament. But on the conceptual side there has been progress. There the crucial step forward was made already before Alexander Bryan Johnson's day, although Johnson did not emulate it. It was made by Bentham in his theory of fictions. Bentham's step was the recognition of contextual definition, or what he called paraphrasis. He recognized that to explain a term we do not need to specify an

object for it to refer to, nor even specify a synonymous word or phrase; we need only show, by whatever means, how to translate all the whole sentences in which the term is to be used. Hume's and Johnson's desperate measure of identifying bodies with impressions ceased to be the only conceivable way of making sense of talk of bodies, even granted that impressions were the only reality. One could undertake to explain talk of bodies in terms of talk of impressions by translating one's whole sentences about bodies into whole sentences about impressions, without equating the bodies themselves to anything at all.

This idea of contextual definition, or recognition of the sentence as the primary vehicle of meaning, was indispensable to the ensuing developments in the foundations of mathematics. It was explicit in Frege, and it attained its full flower in Russell's doctrine of singular descriptions as incomplete symbols.

Contextual definition was one of two resorts that could be expected to have a liberating effect upon the conceptual side of the epistemology of natural knowledge. The other is resort to the resources of set theory as auxiliary concepts. The epistemologist who is willing to eke out his austere ontology of sense impressions with these set-theoretic auxiliaries is suddenly rich: he has not just his impressions to play with, but sets of them, and sets of sets, and so on up. Constructions in the foundations of mathematics have shown that such set-theoretic aids are a powerful addition; after all, the entire glossary of concepts of classical mathematics is constructible from them. Thus equipped, our epistemologist may not need either to identify bodies with impressions or to settle for contextual definition; he may hope to find in some subtle construction of sets upon sets of sense impressions a category of objects enjoying just the formula properties that he wants for bodies.

The two resorts are very unequal in epistemological status. Contextual definition is unassailable. Sentences that have been given meaning as wholes are undeniably meaningful, and the use they make of their component terms is therefore meaningful, regardless of whether any translations are offered for those terms in isolation. Surely Hume and A. B. Johnson would have used contextual definition with pleasure if they had thought of it. Recourse to sets, on the other hand, is a drastic ontological move, a retreat from the austere ontology of impressions. There are philosophers who would rather settle for bodies outright than accept all

these sets, which amount, after all, to the whole abstract ontology of mathematics.

This issue has not always been clear, however, owing to deceptive hints of continuity between elementary logic and set theory. This is why mathematics was once believed to reduce to logic, that is, to an innocent and unquestionable logic, and to inherit these qualities. And this is probably why Russell was content to resort to sets as well as to contextual definition when in *Our Knowledge of the External World* and elsewhere he addressed himself to the epistemology of natural knowledge, on its conceptual side.

To account for the external world as a logical construct of sense data – such, in Russell's terms, was the program. It was Carnap, in his *Der logische Aufbau der Welt* of 1928, who came nearest to executing it.

This was the conceptual side of epistemology; what of the doctrinal? There the Humean predicament remained unaltered. Carnap's constructions, if carried successfully to completion, would have enabled us to translate all sentences about the world into terms of sense data, or observation, plus logic and set theory. But the mere fact that a sentence is *couched* in terms of observation, logic, and set theory does not mean that it can be *proved* from observation sentences by logic and set theory. The most modest of generalizations about observable traits will cover more cases than its utterer can have had occasion actually to observe. The hopelessness of grounding natural science upon immediate experience in a firmly logical way was acknowledged. The Cartesian quest for certainty had been the remote motivation of epistemology, both on its conceptual and its doctrinal side; but that quest was seen as a lost cause. To endow the truths of nature with the full authority of immediate experience was as forlorn a hope as hoping to endow the truths of mathematics with the potential obviousness of elementary logic.

What then could have motivated Carnap's heroic efforts on the conceptual side of epistemology, when hope of certainty on the doctrinal side was abandoned? There were two good reasons still. One was that such constructions could be expected to elicit and clarify the sensory evidence for science, even if the inferential steps between sensory evidence and scientific doctrine must fall short of certainty. The other reason was that such constructions would deepen our understanding of our discourse about the world, even apart from questions of evidence; it would make all cognitive

discourse as clear as observation terms and logic and, I must regretfully add, set theory.

It was sad for epistemologists, Hume and others, to have to acquiesce in the impossibility of strictly deriving the science of the external world from sensory evidence. Two cardinal tenets of empiricism remained unassailable, however, and so remain to this day. One is that whatever evidence there is for science is sensory evidence. The other, to which I shall return, is that all inculcation of meanings of words must rest ultimately on sensory evidence. Hence the continuing attractiveness of the idea of a *logischer Aufbau* in which the sensory content of discourse would stand forth explicitly.

If Carnap had successfully carried such a construction through, how could he have told whether it was the right one? The question would have had no point. He was seeking what he called a *rational reconstruction*. Any construction of physicalistic discourse in terms of sense experience, logic, and set theory would have been seen as satisfactory if it made the physicalistic discourse come out right. If there is one way there are many, but any would be a great achievement.

But why all this creative reconstruction, all this make-believe? The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology? Such a surrender of the epistemological burden to psychology is a move that was disallowed in earlier times as circular reasoning. If the epistemologist's goal is validation of the grounds of empirical science, he defeats his purpose by using psychology or other empirical science in the validation. However, such scruples against circularity have little point once we have stopped dreaming of deducing science from observations. If we are out simply to understand the link between observation and science, we are well advised to use any available information, including that provided by the very science whose link with observation we are seeking to understand.

But there remains a different reason, unconnected with fears of circularity, for still favoring creative reconstruction. We should like to be able to *translate* science into logic and observation terms and set theory. This would be a great epistemological achievement, for it would show all the rest of the concepts of science to be theoretically superfluous. It would legitimize them – to what-

ever degree the concepts of set theory, logic, and observation are themselves legitimate – by showing that everything done with the one apparatus could in principle be done with the other. If psychology itself could deliver a truly translational reduction of this kind, we should welcome it; but certainly it cannot, for certainly we did not grow up learning definitions of physicalistic language in terms of a prior language of set theory, logic, and observation. Here, then, would be good reason for persisting in a rational reconstruction: we want to establish the essential innocence of physical concepts, by showing them to be theoretically dispensable.

The fact is, though, that the construction which Carnap outlined in *Der logische Aufbau der Welt* does not give translational reduction either. It would not even if the outline were filled in. The crucial point comes where Carnap is explaining how to assign sense qualities to positions in physical space and time. These assignments are to be made in such a way as to fulfill, as well as possible, certain desiderata which he states, and with growth of experience the assignments are to be revised to suit. This plan, however illuminating, does not offer any key to *translating* the sentences of science into terms of observation, logic, and set theory.

We must despair of any such reduction. Carnap had despaired of it by 1936, when, in "Testability and Meaning,"² he introduced so-called *reduction forms* of a type weaker than definition. Definitions had shown always how to translate sentences into equivalent sentences. Contextual definition of a term showed how to translate sentences containing the term into equivalent sentences lacking the term. Reduction forms of Carnap's liberalized kind, on the other hand, do not in general give equivalences; they give implications. They explain a new term, if only partially, by specifying some sentences which are implied by sentences containing the term, and other sentences which imply sentences containing the term.

It is tempting to suppose that the countenancing of reduction forms in this liberal sense is just one further step of liberalization comparable to the earlier one, taken by Bentham, of countenancing contextual definition. The former and sterner kind of rational reconstruction might have been represented as a fictitious history in which we imagined our ancestors introducing the terms of physicalistic discourse on a phenomenistic and set-theoretic basis by a succession of contextual definitions. The new and more liberal kind of rational recon-

struction is a fictitious history in which we imagine our ancestors introducing those terms by a succession rather of reduction forms of the weaker sort.

This, however, is a wrong comparison. The fact is rather that the former and sterner kind of rational reconstruction, where definition reigned, embodied no fictitious history at all. It was nothing more nor less than a set of directions – or would have been, if successful – for accomplishing everything in terms of phenomena and set theory that we now accomplish in terms of bodies. It would have been a true reduction by translation, a legitimation by elimination. *Definire est eliminare*. Rational reconstruction by Carnap's later and looser reduction forms does none of this.

To relax the demand for definition, and settle for a kind of reduction that does not eliminate, is to renounce the last remaining advantage that we supposed rational reconstruction to have over straight psychology; namely, the advantage of translational reduction. If all we hope for is a reconstruction that links science to experience in explicit ways short of translation, then it would seem more sensible to settle for psychology. Better to discover how science is in fact developed and learned than to fabricate a fictitious structure to a similar effect.

The empiricist made one major concession when he despaired of deducing the truths of nature from sensory evidence. In despairing now even of translating those truths into terms of observation and logico-mathematical auxiliaries, he makes another major concession. For suppose we hold, with the old empiricist Peirce, that the very meaning of a statement consists in the difference its truth would make to possible experience. Might we not formulate, in a chapter-length sentence in observational language, all the difference that the truth of a given statement might make to experience, and might we not then take all this as the translation? Even if the difference that the truth of the statement would make to experience ramifies indefinitely, we might still hope to embrace it all in the logical implications of our chapter-length formulation, just as we can axiomatize an infinity of theorems. In giving up hope of such translation, then, the empiricist is conceding that the empirical meanings of typical statements about the external world are inaccessible and ineffable.

How is this inaccessibility to be explained? Simply on the ground that the experiential implications of a typical statement about bodies are too

complex for finite axiomatization, however lengthy? No; I have a different explanation. It is that the typical statement about bodies has no fund of experiential implications it can call its own. A substantial mass of theory, taken together, will commonly have experiential implications; this is how we make verifiable predictions. We may not be able to explain why we arrive at theories which make successful predictions, but we do arrive at such theories.

Sometimes also an experience implied by a theory fails to come off; and then, ideally, we declare the theory false. But the failure falsifies only a block of theory as a whole, a conjunction of many statements. The failure shows that one or more of those statements is false, but it does not show which. The predicted experiences, true and false, are not implied by any one of the component statements of the theory rather than another. The component statements simply do not have empirical meanings, by Peirce's standard, but a sufficiently inclusive portion of theory does. If we can aspire to a sort of *logischer Aufbau der Welt* at all, it must be to one in which the texts slated for translation into observational and logico-mathematical terms are mostly broad theories taken as wholes, rather than just terms or short sentences. The translation of a theory would be a ponderous axiomatization of all the experiential difference that the truth of the theory would make. It would be a queer translation, for it would translate the whole but none of the parts. We might better speak in such a case not of translation but simply of observational evidence for theories; and we may, following Peirce, still fairly call this the empirical meaning of the theories.

These considerations raise a philosophical question even about ordinary unphilosophical translation, such as from English into Arunta or Chinese. For, if the English sentences of a theory have their meaning only together as a body, then we can justify their translation into Arunta only together as a body. There will be no justification for pairing off the component English sentences with component Arunta sentences, except as these correlations make the translation of the theory as a whole come out right. Any translations of the English sentences into Arunta sentences will be as correct as any other, so long as the net empirical implications of the theory as a whole are preserved in translation. But it is to be expected that many different ways of translating the component sentences, essentially different individually, would

deliver the same empirical implications for the theory as a whole; deviations in the translation of one component sentence could be compensated for in the translation of another component sentence. Insofar, there can be no ground for saying which of two glaringly unlike translations of individual sentences is right.³

For an uncritical mentalist, no such indeterminacy threatens. Every term and every sentence is a label attached to an idea, simple or complex, which is stored in the mind. When on the other hand we take a verification theory of meaning seriously, the indeterminacy would appear to be inescapable. The Vienna Circle espoused a verification theory of meaning but did not take it seriously enough. If we recognize with Peirce that the meaning of a sentence turns purely on what would count as evidence for its truth, and if we recognize with Duhem that theoretical sentences have their evidence not as single sentences but only as larger blocks of theory, then the indeterminacy of translation of theoretical sentences is the natural conclusion. And most sentences, apart from observation sentences, are theoretical. This conclusion, conversely, once it is embraced, seals the fate of any general notion of propositional meaning or, for that matter, state of affairs.

Should the unwelcomeness of the conclusion persuade us to abandon the verification theory of meaning? Certainly not. The sort of meaning that is basic to translation, and to the learning of one's own language, is necessarily empirical meaning and nothing more. A child learns his first words and sentences by hearing and using them in the presence of appropriate stimuli. These must be external stimuli, for they must act both on the child and on the speaker from whom he is learning.⁴ Language is socially inculcated and controlled; the inculcation and control turn strictly on the keying of sentences to shared stimulation. Internal factors may vary *ad libitum* without prejudice to communication as long as the keying of language to external stimuli is undisturbed. Surely one has no choice but to be an empiricist so far as one's theory of linguistic meaning is concerned.

What I have said of infant learning applies equally to the linguist's learning of a new language in the field. If the linguist does not lean on related languages for which there are previously accepted translation practices, then obviously he had no data but the concomitances of native utterance and observable stimulus situation. No wonder there is indeterminacy of translation – for of

course only a small fraction of our utterances report concurrent external stimulation. Granted, the linguist will end up with unequivocal translations of everything; but only by making many arbitrary choices – arbitrary even though unconscious – along the way. Arbitrary? By this I mean that different choices could still have made everything come out right that is susceptible in principle to any kind of check.

Let me link up, in a different order, some of the points I have made. The crucial consideration behind my argument for the indeterminacy of translation was that a statement about the world does not always or usually have a separable fund of empirical consequences that it can call its own. That consideration served also to account for the impossibility of an epistemological reduction of the sort where every sentence is equated to a sentence in observational and logico-mathematical terms. And the impossibility of that sort of epistemological reduction dissipated the last advantage that rational reconstruction seemed to have over psychology.

Philosophers have rightly despaired of translating everything into observational and logico-mathematical terms. They have despaired of this even when they have not recognized, as the reason for this irreducibility, that the statements largely do not have their private bundles of empirical consequences. And some philosophers have seen in this irreducibility the bankruptcy of epistemology. Carnap and the other logical positivists of the Vienna Circle had already pressed the term “*metaphysics*” into pejorative use, as connoting meaninglessness; and the term “*epistemology*” was next. Wittgenstein and his followers, mainly at Oxford, found a residual philosophical vocation in therapy: in curing philosophers of the delusion that there were epistemological problems.

But I think that at this point it may be more useful to say rather that epistemology still goes on, though in a new setting and a clarified status. Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input – certain patterns of irradiation in assorted frequencies, for instance – and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation

that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one’s theory of nature transcends any available evidence.

Such a study could still include, even, something like the old rational reconstruction, to whatever degree such reconstruction is practicable; for imaginative constructions can afford hints of actual psychological processes, in much the way that mechanical simulations can. But a conspicuous difference between old epistemology and the epistemological enterprise in this new psychological setting is that we can now make free use of empirical psychology.

The old epistemology aspired to contain, in a sense, natural science; it would construct it somehow from sense data. Epistemology in its new setting, conversely, is contained in natural science, as a chapter of psychology. But the old containment remains valid too, in its way. We are studying how the human subject of our study posits bodies and projects his physics from his data, and we appreciate that our position in the world is just like his. Our very epistemological enterprise, therefore, and the psychology wherein it is a component chapter, and the whole of natural science wherein psychology is a component book – all this is our own construction or projection from stimulations like those we were meting out to our epistemological subject. There is thus reciprocal containment, though containment in different senses: epistemology in natural science and natural science in epistemology.

This interplay is reminiscent again of the old threat of circularity, but it is all right now that we have stopped dreaming of deducing science from sense data. We are after an understanding of science as an institution or process in the world, and we do not intend that understanding to be any better than the science which is its object. This attitude is indeed one that Neurath was already urging in Vienna Circle days, with his parable of the mariner who has to rebuild his boat while staying afloat in it.

One effect of seeing epistemology in a psychological setting is that it resolves a stubborn old enigma of epistemological priority. Our retinas are irradiated in two dimensions, yet we see things as three-dimensional without conscious inference. Which is to count as observation – the unconscious two-dimensional reception or the conscious three-dimensional apprehension? In the old epistemology

gical context the conscious form had priority, for we were out to justify our knowledge of the external world by rational reconstruction, and that demands awareness. Awareness ceased to be demanded when we gave up trying to justify our knowledge of the external world by rational reconstruction. What to count as observation now can be settled in terms of the stimulation of sensory receptors, let consciousness fall where it may.

The Gestalt psychologists' challenge to sensory atomism, which seemed so relevant to epistemology forty years ago, is likewise deactivated. Regardless of whether sensory atoms or Gestalten are what favor the forefront of our consciousness, it is simply the stimulations of our sensory receptors that are best looked upon as the input to our cognitive mechanism. Old paradoxes about unconscious data and inference, old problems about chains of inference that would have to be completed too quickly – these no longer matter.

In the old anti-psychologistic days the question of epistemological priority was moot. What is epistemologically prior to what? Are Gestalten prior to sensory atoms because they are noticed, or should we favor sensory atoms on some more subtle ground? Now that we are permitted to appeal to physical stimulation, the problem dissolves; *A* is epistemologically prior to *B* if *A* is causally nearer than *B* to the sensory receptors. Or, what is in some ways better, just talk explicitly in terms of causal proximity to sensory receptors and drop the talk of epistemological priority.

Around 1932 there was debate in the Vienna Circle over what to count as observation sentences, or *Protokollsätze*.⁵ One position was that they had the form of reports of sense impressions. Another was that they were statements of an elementary sort about the external world, e.g., "A red cube is standing on the table." Another, Neurath's, was that they had the form of reports of relations between percipients and external things: "Otto now sees a red cube on the table." The worst of it was that there seemed to be no objective way of settling the matter: no way of making real sense of the question.

Let us now try to view the matter unreservedly in the context of the external world. Vaguely speaking, what we want of observation sentences is that they be the ones in closest causal proximity to the sensory receptors. But how is such proximity to be gauged? The idea may be rephrased this way: observation sentences are sentences which, as we learn language, are most strongly conditioned

to concurrent sensory stimulation rather than to stored collateral information. Thus let us imagine a sentence queried for our verdict as to whether it is true or false, queried for our assent or dissent. Then the sentence is an observation sentence if our verdict depends only on the sensory stimulation present at the time.

But a verdict cannot depend on present stimulation to the exclusion of stored information. The very fact of our having learned the language evinces much storing of information, and of information without which we should be in no position to give verdicts on sentences however observational. Evidently then we must relax our definition of observation sentence to read thus: a sentence is an observation sentence if all verdicts on it depend on present sensory stimulation and on no stored information beyond what goes into understanding the sentence.

This formulation raises another problem: how are we to distinguish between information that goes into understanding a sentence and information that goes beyond? This is the problem of distinguishing between analytic truth, which issues from the mere meanings of words, and synthetic truth, which depends on more than meanings. Now I have long maintained that this distinction is illusory. There is one step toward such a distinction, however, which does make sense: a sentence that is true by mere meanings of words should be expected, at least if it is simple, to be subscribed to by all fluent speakers in the community. Perhaps the controversial notion of analyticity can be dispensed with, in our definition of observation sentence, in favor of this straightforward attribute of community-wide acceptance.

This attribute is of course no explication of analyticity. The community would agree that there have been black dogs, yet none who talk of analyticity would call this analytic. My rejection of the analyticity notion just means drawing no line between what goes into the mere understanding of the sentences of a language and what else the community sees eye-to-eye on. I doubt that an objective distinction can be made between meaning and such collateral information as is community-wide.

Turning back then to our task of defining observation sentences, we get this: an observation sentence is one on which all speakers of the language give the same verdict when given the same concurrent stimulation. To put the point negatively, an observation sentence is one that is not sensitive

to differences in past experience within the speech community.

This formulation accords perfectly with the traditional role of the observation sentence as the court of appeal of scientific theories. For by our definition the observation sentences are the sentences on which all members of the community will agree under uniform stimulation. And what is the criterion of membership in the same community? Simply, general fluency of dialogue. This criterion admits of degrees, and indeed we may usefully take the community more narrowly for some studies than for others. What count as observation sentences for a community of specialists would not always so count for a larger community.

There is generally no subjectivity in the phrasing of observation sentences, as we are now conceiving them; they will usually be about bodies. Since the distinguishing trait of an observation sentence is intersubjective agreement under agreeing stimulation, a corporeal subject matter is likelier than not.

The old tendency to associate observation sentences with a subjective sensory subject matter is rather an irony when we reflect that observation sentences are also meant to be the intersubjective tribunal of scientific hypotheses. The old tendency was due to the drive to base science on something firmer and prior in the subject's experience; but we dropped that project.

The dislodging of epistemology from its old status of first philosophy loosed a wave, we saw, of epistemological nihilism. This mood is reflected somewhat in the tendency of Polányi, Kuhn, and the late Russell Hanson to belittle the role of evidence and to accentuate cultural relativism. Hanson ventured even to discredit the idea of observation, arguing that so-called observations vary from observer to observer with the amount of knowledge that the observers bring with them. The veteran physicist looks at some apparatus and sees an x-ray tube. The neophyte, looking at the same place, observes rather "a glass and metal instrument replete with wires, reflectors, screws, lamps, and pushbuttons."⁶ One man's observation is another man's closed book or flight of fancy. The notion of observation as the impartial and objective source of evidence for science is bankrupt. Now my answer to the x-ray example was already hinted a little while back: what counts as an observation sentence varies with the width of community considered. But we can also always get an absolute standard by taking in all speakers of the

language, or most.⁷ It is ironic that philosophers, finding the old epistemology untenable as a whole, should react by repudiating a part which has only now moved into clear focus.

Clarification of the notion of observation sentence is a good thing, for the notion is fundamental in two connections. These two correspond to the duality that I remarked upon early in this essay: the duality between concept and doctrine, between knowing what a sentence means and knowing whether it is true. The observation sentence is basic to both enterprises. Its relation to doctrine, to our knowledge of what is true, is very much the traditional one: observation sentences are the repository of evidence for scientific hypotheses. Its relation to meaning is fundamental too, since observation sentences are the ones we are in a position to learn to understand first, both as children and as field linguists. For observation sentences are precisely the ones that we can correlate with observable circumstances of the occasion of utterance or assent, independently of variations in the past histories of individual informants. They afford the only entry to a language.

The observation sentence is the cornerstone of semantics. For it is, as we just saw, fundamental to the learning of meaning. Also, it is where meaning is firmest. Sentences higher up in theories have no empirical consequences they can call their own; they confront the tribunal of sensory evidence only in more or less inclusive aggregates. The observation sentence, situated at the sensory periphery of the body scientific, is the minimal verifiable aggregate; it has an empirical content all its own and wears it on its sleeve.

The predicament of the indeterminacy of translation has little bearing on observation sentences. The equating of an observation sentence of our language to an observation sentence of another language is mostly a matter of empirical generalization; it is a matter of identity between the range of stimulations that would prompt assent to the one sentence and the range of stimulations that would prompt assent to the other.⁸

It is no shock to the preconceptions of old Vienna to say that epistemology now becomes semantics. For epistemology remains centered as always on evidence, and meaning remains centered as always on verification; and evidence is verification. What is likelier to shock preconceptions is that meaning, once we get beyond observation sentences, ceases in general to have any clear applicability to single sentences; also that episte-

mology merges with psychology, as well as with linguistics.

This rubbing out of boundaries could contribute to progress, it seems to me, in philosophically interesting inquiries of a scientific nature. One possible area is perceptual norms. Consider, to begin with, the linguistic phenomenon of phonemes. We form the habit, in hearing the myriad variations of spoken sounds, of treating each as an approximation to one or another of a limited number of norms – around thirty altogether – constituting so to speak a spoken alphabet. All speech in our language can be treated in practice as sequences of just those thirty elements, thus rectifying small deviations. Now outside the realm of language also there is probably only a rather lim-

ited alphabet of perceptual norms altogether, toward which we tend unconsciously to rectify all perceptions. These, if experimentally identified, could be taken as epistemological building blocks, the working elements of experience. They might prove in part to be culturally variable, as phonemes are, and in part universal.

Again there is the area that the psychologist Donald T. Campbell calls evolutionary epistemology.⁹ In this area there is work by Hüseyin Yilmaz, who shows how some structural traits of color perception could have been predicted from survival value.¹⁰ And a more emphatically epistemological topic that evolution helps to clarify is induction, now that we are allowing epistemology the resources of natural science.¹¹

Notes

- 1 A. B. Johnson, *A Treatise on Language* (New York, 1836; Berkeley, 1947).
- 2 Carnap, *Philosophy of Science* 3 (1936), pp. 419–71; 4 (1937), pp. 1–40.
- 3 See Quine, *Ontological Relativity* (New York: Columbia University Press, 1969), pp. 2ff.
- 4 See *ibid.*, p. 28.
- 5 Carnap and Neurath in *Erkenntnis* 3 (1932), pp. 204–28.
- 6 N. R. Hanson, "Observation and Interpretation," in S. Morgenbesser (ed.), *Philosophy of Science Today* (New York: Basic Books, 1966).
- 7 This qualification allows for occasional deviants such as the insane or the blind. Alternatively, such cases might be excluded by adjusting the level of fluency of dialogue whereby we define sameness of language.

(For prompting this note and influencing the development of this essay also in more substantial ways I am indebted to Burton Dreben.)

- 8 Cf. Quine, *Word and Object* (Cambridge, MA: MIT Press, 1960), pp. 31–46, 68.
- 9 D. T. Campbell, "Methodological Suggestions from a Comparative Psychology of Knowledge Processes," *Inquiry* 2 (1959), pp. 152–82.
- 10 Hüseyin Yilmaz, "On Color Vision and a New Approach to General Perception," in E. E. Bernard and M. R. Kare (eds), *Biological Prototypes and Synthetic Systems* (New York: Plenum, 1962); "Perceptual Invariance and the Psychophysical Law," *Perception and Psychophysics* 2 (1967), pp. 533–8.
- 11 See Quine, "Natural Kinds," in *Ontological Relativity*, ch. 5.

What Is "Naturalized Epistemology"?

Jaegwon Kim

Epistemology as a Normative Inquiry

Descartes's epistemological inquiry in the *Meditations* begins with this question: What propositions are worthy of belief? In the First Meditation Descartes canvasses beliefs of various kinds he had formerly held as true and finds himself forced to conclude that he ought to reject them, that he ought not to accept them as true. We can view Cartesian epistemology as consisting of the following two projects: to identify the criteria by which we ought to regulate acceptance and rejection of beliefs, and to determine what we may be said to know according to those criteria. Descartes's epistemological agenda has been the agenda of Western epistemology to this day. The twin problems of identifying criteria of justified belief and coming to terms with the skeptical challenge to the possibility of knowledge have defined the central tasks of theory of knowledge since Descartes. This was as true of the empiricists, of Locke and Hume and Mill, as of those who more closely followed Descartes in the rationalist path.¹

It is no wonder then that modern epistemology has been dominated by a single concept, that of *justification*, and two fundamental questions involving it: What conditions must a belief meet if we are justified in accepting it as true? and What beliefs are we in fact justified in accepting? Note that the first question does not ask for an "analysis" or "meaning" of the term "justified belief." And it is generally assumed, even if not always

explicitly stated, that not just any statement of a necessary and sufficient condition for a belief to be justified will do. The implicit requirement has been that the stated conditions must constitute "criteria" of justified belief, and for this it is necessary that the conditions be stated *without the use of epistemic terms*. Thus, formulating conditions of justified belief in such terms as "adequate evidence," "sufficient ground," "good reason," "beyond a reasonable doubt," and so on, would be merely to issue a promissory note redeemable only when these epistemic terms are themselves explained in a way that accords with the requirement.²

This requirement, while it points in the right direction, does not go far enough. What is crucial is this: *the criteria of justified belief must be formulated on the basis of descriptive or naturalistic terms alone, without the use of any evaluative or normative ones, whether epistemic or of another kind.*³ Thus, an analysis of justified belief that makes use of such terms as "intellectual requirement"⁴ and "having a right to be sure"⁵ would not satisfy this generalized condition; although such an analysis can be informative and enlightening about the interrelationships of these normative concepts, it will not, on the present conception, count as a statement of *criteria* of justified belief, unless of course these terms are themselves provided with nonnormative criteria. What is problematic, therefore, about the use of epistemic terms in stating criteria of justified belief is not its possible circularity in the usual sense; rather it is the fact that these epistemic terms are themselves essentially normative. We shall later discuss the rationale of this strengthened requirement.

Originally published in J. Tomberlin (ed.), *Philosophical Perspectives*. 2. *Epistemology* (Atascadero, CA: Ridgeview Publishing Co., 1988).

As many philosophers have observed,⁶ the two questions we have set forth, one about the criteria of justified belief and the other about what we can be said to know according to those criteria, constrain each other. Although some philosophers have been willing to swallow skepticism just because what we regard as correct criteria of justified belief are seen to lead inexorably to the conclusion that none, or very few, of our beliefs are justified, the usual presumption is that our answer to the first question should leave our epistemic situation largely unchanged. That is to say, it is expected to turn out that according to the criteria of justified belief we come to accept, we know, or are justified in believing, pretty much what we reflectively think we know or are entitled to believe.

Whatever the exact history, it is evident that the concept of justification has come to take center stage in our reflections on the nature of knowledge. And apart from history, there is a simple reason for our preoccupation with justification: it is the only specifically epistemic component in the classic tripartite conception of knowledge. Neither belief nor truth is a specifically epistemic notion: belief is a psychological concept and truth a semantical-metaphysical one. These concepts may have an implicit epistemological dimension, but if they do, it is likely to be through their involvement with essentially normative epistemic notions like justification, evidence, and rationality. Moreover, justification is what makes knowledge itself a normative concept. On the surface at least, neither truth nor belief is normative or evaluative (I shall argue below, though, that belief does have an essential normative dimension). But justification manifestly is normative. If a belief is justified for us, then it is *permissible* and *reasonable*, from the epistemic point of view, for us to hold it, and it would be *epistemically irresponsible* to hold beliefs that contradict it. If we consider believing or accepting a proposition to be an "action" in an appropriate sense, belief justification would then be a special case of justification of action, which in its broadest terms is the central concern of normative ethics. Just as it is the business of normative ethics to delineate the conditions under which acts and decisions are justified from the moral point of view, so it is the business of epistemology to identify and analyze the conditions under which beliefs, and perhaps other propositional attitudes, are justified from the epistemological point of view. It probably is only an historical accident

that we standardly speak of "normative ethics" but not of "normative epistemology." Epistemology is a normative discipline as much as, and in the same sense as, normative ethics.

We can summarize our discussion thus far in the following points: that justification is a central concept of our epistemological tradition, that justification, as it is understood in this tradition, is a normative concept, and in consequence that epistemology itself is a normative inquiry whose principal aim is a systematic study of the conditions of justified belief. I take it that these points are uncontroversial, although of course there could be disagreement about the details – for example, about what it means to say a concept or theory is "normative" or "evaluative."

The Foundationalist Strategy

In order to identify the target of the naturalistic critique – in particular, Quine's – it will be useful to take a brief look at the classic response to the epistemological program set forth by Descartes. Descartes's approach to the problem of justification is a familiar story, at least as the textbook tells it: it takes the form of what is now commonly called "foundationalism." The foundationalist strategy is to divide the task of explaining justification into two stages: first, to identify a set of beliefs that are "directly" justified in that they are justified without deriving their justified status from that of any other belief, and then to explain how other beliefs may be "indirectly" or "inferentially" justified by standing in an appropriate relation to those already justified. Directly justified beliefs, or "basic beliefs," are to constitute the foundation upon which the superstructure of "nonbasic" or "derived" beliefs is to rest. What beliefs then are directly justified, according to Descartes? Subtleties aside, he claimed that beliefs about our own present conscious states are among them. In what does their justification consist? What is it about these beliefs that makes them directly justified? Somewhat simplistically again, Descartes's answer is that they are justified because they are *indubitable*, that the attentive and reflective mind *cannot but assent* to them. How are nonbasic beliefs justified? By "deduction" – that is, by a series of inferential steps, or "intuitions," each of which is indubitable. If, therefore, we take Cartesian indubitability as a psychological notion, Descartes's epistemological theory can be said to meet the

desideratum of providing nonepistemic, naturalistic criteria of justified belief.

Descartes's foundationalist program was inherited, in its essential outlines, by the empiricists. In particular, his "mentalism," that beliefs about one's own current mental state are epistemologically basic, went essentially unchallenged by the empiricists and positivists, until this century. Epistemologists have differed from one another chiefly in regard to two questions: first, what else belonged in our corpus of basic beliefs, and second, how the derivation of the nonbasic part of our knowledge was to proceed. Even the Logical Positivists were, by and large, foundationalists, although some of them came to renounce Cartesian mentalism in favor of a "physicalistic basis."⁷ In fact, the Positivists were foundationalists twice over: for them "observation," whether phenomenological or physical, served not only as the foundation of knowledge but as the foundation of all "cognitive meaning" – that is, as both an epistemological and a semantic foundation.

Quine's Arguments

It has become customary for epistemologists who profess allegiance to a "naturalistic" conception of knowledge to pay homage to Quine as the chief contemporary provenance of their inspiration – especially to his influential paper "Epistemology Naturalized."⁸ Quine's principal argument in this paper against traditional epistemology is based on the claim that the Cartesian foundationalist program has failed – that the Cartesian "quest for certainty" is "a lost cause." While this claim about the hopelessness of the Cartesian "quest for certainty" is nothing new, using it to discredit the very conception of normative epistemology is new, something that any serious student of epistemology must contend with.

Quine divides the classic epistemological program into two parts: *conceptual reduction* whereby physical terms, including those of theoretical science, are reduced, via definition, to terms referring to phenomenal features of sensory experience, and *doctrinal reduction* whereby truths about the physical world are appropriately obtained from truths about sensory experience. The "appropriateness" just alluded to refers to the requirement that the favored epistemic status ("certainty" for classic epistemologists, according to Quine) of our basic beliefs be transferred, essentially undimin-

ished, to derived beliefs, a necessary requirement if the derivational process is to yield knowledge from knowledge. What derivational methods have this property of preserving epistemic status? Perhaps there are none, given our proneness to err in framing derivations as in anything else, not to mention the possibility of lapses of attention and memory in following lengthy proofs. But logical deduction comes as close to being one as any; it can at least be relied on to transmit truth, if not epistemic status. It could perhaps be argued that no method can preserve certainty unless it preserves (or is known to preserve) truth; and if this is so, logical deduction is the only method worth considering. I do not know whether this was the attitude of most classic epistemologists; but Quine assumes that if deduction doesn't fill their bill, nothing will.

Quine sees the project of conceptual reduction as culminating in Carnap's *Der logische Aufbau der Welt*. As Quine sees it, Carnap "came nearest to executing" the conceptual half of the classic epistemological project. But coming close is not good enough. Because of the holistic manner in which empirical meaning is generated by experience, no reduction of the sort Carnap and others so eagerly sought could in principle be completed. For definitional reduction requires point-to-point meaning relations⁹ between physical terms and phenomenal terms, something that Quine's holism tells us cannot be had. The second half of the program, doctrinal reduction, is in no better shape; in fact, it was the one to stumble first, for, according to Quine, its impossibility was decisively demonstrated long before the *Aufbau*, by Hume in his celebrated discussion of induction. The "Humean predicament" shows that theory cannot be logically deduced from observation; there simply is no way of deriving theory from observation that will transmit the latter's epistemic status intact to the former.

I don't think anyone wants to disagree with Quine in these claims. It is not possible to "validate" science on the basis of sensory experience, if "validation" means justification through logical deduction. Quine of course does not deny that our theories depend on observation for evidential support; he has said that sensory evidence is the only evidence there is. To be sure, Quine's argument against the possibility of conceptual reduction has a new twist: the application of his "holism." But his conclusion is no surprise; "translational phenomenalism" has been

moribund for many years.¹⁰ And, as Quine himself notes, his argument against the doctrinal reduction, the “quest for certainty,” is only a restatement of Hume’s “skeptical” conclusions concerning induction: induction after all is not deduction. Most of us are inclined, I think, to view the situation Quine describes with no great alarm, and I rather doubt that these conclusions of Quine’s came as news to most epistemologists when “Epistemology Naturalized” was first published. We are tempted to respond: of course we can’t define physical concepts in terms of sense-data; of course observation “underdetermines” theory. That is why observation is observation and not theory.

So it is agreed on all hands that the classical epistemological project, conceived as one of deductively validating physical knowledge from indubitable sensory data, cannot succeed. But what is the moral of this failure? What should be its philosophical lesson to us? Having noted the failure of the Cartesian program, Quine goes on:¹¹

The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology? Such a surrender of the epistemological burden to psychology is a move that was disallowed in earlier times as circular reasoning. If the epistemologist’s goal is validation of the grounds of empirical science, he defeats his purpose by using psychology or other empirical science in the validation. However, such scruples against circularity have little point once we have stopped dreaming of deducing science from observation. If we are out simply to understand the link between observation and science, we are well advised to use any available information, including that provided by the very science whose link with observation we are seeking to understand.

And Quine has the following to say about the failure of Carnap’s reductive program in the *Aufbau*:¹²

To relax the demand for definition, and settle for a kind of reduction that does not eliminate, is to renounce the last remaining advantage that we supposed rational reconstruction to have over straight psychology; namely, the advantage

of translational reduction. If all we hope for is a reconstruction that links science to experience in explicit ways short of translation, then it would seem more sensible to settle for psychology. Better to discover how science is in fact developed and learned than to fabricate a fictitious structure to a similar effect.

If a task is entirely hopeless, if we know it cannot be executed, no doubt it is rational to abandon it; we would be better off doing something else that has some hope of success. We can agree with Quine that the “validation” – that is, logical deduction – of science on the basis of observation cannot be had; so it is rational to abandon this particular epistemological program, if indeed it ever was a program that anyone seriously undertook. But Quine’s recommendations go further. In particular, there are two aspects of Quine’s proposals that are of special interest to us: first, he is not only advising us to quit the program of “validating science,” but urging us to take up another specific project, an empirical psychological study of our cognitive processes; second, he is also claiming that this new program replaces the old, that both programs are part of something appropriately called “epistemology.” Naturalized epistemology is to be a kind of epistemology after all, a “successor subject”¹³ to classical epistemology.

How should we react to Quine’s urgings? What should be our response? The Cartesian project of validating science starting from the indubitable foundation of first-person psychological reports (perhaps with the help of certain indubitable first principles) is not the whole of classical epistemology – or so it would seem at first blush. In our characterization of classical epistemology, the Cartesian program was seen as one possible response to the problem of epistemic justification, the two-part project of identifying the criteria of epistemic justification and determining what beliefs are in fact justified according to those criteria. In urging “naturalized epistemology” on us, Quine is not suggesting that we give up the Cartesian foundationalist solution and explore others within the same framework¹⁴ – perhaps, to adopt some sort of “coherentist” strategy, or to require of our basic beliefs only some degree of “initial credibility” rather than Cartesian certainty, or to permit some sort of probabilistic derivation in addition to deductive derivation of nonbasic knowledge, or to consider the use of special rules of evidence, like Chisholm’s “principles of evidence,”¹⁵ or to give

up the search for a derivational process that transmits undiminished certainty in favor of one that can transmit diminished but still useful degrees of justification. Quine's proposal is more radical than that. He is asking us to set aside the entire framework of justification-centered epistemology. That is what is new in Quine's proposals. Quine is asking us to put in its place a purely descriptive, causal-nomological science of human cognition.¹⁶

How should we characterize in general terms the difference between traditional epistemological programs, such as foundationalism and coherence theory, on the one hand and Quine's program of naturalized epistemology on the other? Quine's stress is on the *factual* and *descriptive* character of his program; he says, "Why not see how [the construction of theory from observation] *actually proceeds*? Why not settle for psychology?"¹⁷ again, "Better to *discover how science is in fact developed and learned than . . .*"¹⁸ We are given to understand that in contrast traditional epistemology is not a descriptive, factual inquiry. Rather, it is an attempt at a "validation" or "rational reconstruction" of science. Validation, according to Quine, proceeds via deduction, and rational reconstruction via definition. However, their *point* is justificatory – that is, to rationalize our sundry knowledge claims. So Quine is asking us to set aside what is "rational" in rational reconstruction.

Thus, it is normativity that Quine is asking us to repudiate. Although Quine does not explicitly characterize traditional epistemology as "normative" or "prescriptive," his meaning is unmistakable. Epistemology is to be "a chapter of psychology," a law-based predictive-explanatory theory, like any other theory within empirical science; its principal job is to see how human cognizers develop theories (their "picture of the world") from observation ("the stimulation of their sensory receptors"). Epistemology is to go out of the business of justification. We earlier characterized traditional epistemology as essentially normative; we see why Quine wants us to reject it. Quine is urging us to replace a normative theory of cognition with a descriptive science.

Losing Knowledge from Epistemology

If justification drops out of epistemology, knowledge itself drops out of epistemology. For our concept of knowledge is inseparably tied to that

of justification. As earlier noted, knowledge itself is a normative notion. Quine's nonnormative, naturalized epistemology has no room for our concept of knowledge. It is not surprising that, in describing naturalized epistemology, Quine seldom talks about knowledge; instead, he talks about "science" and "theories" and "representations." Quine would have us investigate how sensory stimulation "leads" to "theories" and "representation" of the world. I take it that within the traditional scheme these "theories" and "representations" correspond to beliefs, or systems of beliefs; thus, what Quine would have us do is to investigate how sensory stimulation leads to the formation of beliefs about the world.

But in what sense of "lead"? I take it that Quine has in mind a causal or nomological sense. He is urging us to develop a theory, an empirical theory, that uncovers lawful regularities governing the processes through which organisms come to develop beliefs about their environment as a causal result of having their sensory receptors stimulated in certain ways. Quine says:¹⁹

[Naturalized epistemology] studies a natural phenomenon, viz., a physical human subject. This human subject is accorded experimentally controlled input – certain patterns of irradiation in assorted frequencies, for instance – and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. *The relation between the meager input and torrential output* is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see *how evidence relates to theory*, and in what ways one's theory of nature transcends any available evidence.

The relation Quine speaks of between "meager input" and "torrential output" is a causal relation; at least it is qua causal relation that the naturalized epistemologist investigates it. It is none of the naturalized epistemologist's business to assess whether, and to what degree, the input "justifies" the output, how a given irradiation of the subject's retinas makes it "reasonable" or "rational" for the subject to emit certain representational output. His interest is strictly causal and nomological: he wants us to look for patterns of lawlike dependencies characterizing the input-output relations for this particular organism and others of a like physical structure.

If this is right, it makes Quine's attempt to relate his naturalized epistemology to traditional epistemology look at best lame. For in what sense is the study of causal relationships between physical stimulation of sensory receptors and the resulting cognitive output a way of "seeing how evidence relates to theory" in an epistemologically relevant sense? The causal relation between sensory input and cognitive output is a relation between "evidence" and "theory"; however, it is not an *evidential relation*. This can be seen from the following consideration: the nomological patterns that Quine urges us to look for are certain to vary from species to species, depending on the particular way each biological (and possibly nonbiological) species processes information, but the evidential relation in its proper normative sense must abstract from such factors and concern itself only with the degree to which evidence supports hypothesis.

In any event, the concept of evidence is inseparable from that of justification. When we talk of "evidence" in an epistemological sense we are talking about justification: one thing is "evidence" for another just in case the first tends to enhance the reasonableness or justification of the second. And such evidential relations hold in part because of the "contents" of the items involved, not merely because of the causal or nomological connections between them. A strictly nonnormative concept of evidence is not our concept of evidence; it is something that we do not understand.²⁰

None of us, I think, would want to quarrel with Quine about the interest or importance of the psychological study of how our sensory input causes our epistemic output. This is only to say that the study of human (or other kinds of) cognition is of interest. That isn't our difficulty; our difficulty is whether, and in what sense, pursuing Quine's "epistemology" is a way of doing epistemology – that is, a way of studying "how evidence relates to theory." Perhaps, Quine's recommendation that we discard justification-centered epistemology is worth pondering; and his exhortation to take up the study of psychology perhaps deserves to be heeded also. What is mysterious is why this recommendation has to be coupled with the rejection of normative epistemology (if normative epistemology is not a possible inquiry, why shouldn't the would-be epistemologist turn to, say, hydrodynamics or ornithology rather than psychology?). But of course Quine is saying more; he is saying that an understandable, if misguided, motivation (that is, seeing "how evidence relates to theory")

does underlie our proclivities for indulgence in normative epistemology, but that we would be better served by a scientific study of human cognition than normative epistemology.

But it is difficult to see how an "epistemology" that has been purged of normativity, one that lacks an appropriate normative concept of justification or evidence, can have anything to do with the concerns of traditional epistemology. And unless naturalized epistemology and classical epistemology share some of their central concerns, it's difficult to see how one could *replace* the other, or be a way (a better way) of doing the other.²¹ To be sure, they both investigate "how evidence relates to theory." But putting the matter this way can be misleading, and has perhaps misled Quine: the two disciplines do not investigate the same relation. As lately noted, normative epistemology is concerned with the evidential relation properly so-called – that is, the relation of justification – and Quine's naturalized epistemology is meant to study the causal – nomological relation. For epistemology to go out of the business of justification is for it to go out of business.

Belief Attribution and Rationality

Perhaps we have said enough to persuade ourselves that Quine's naturalized epistemology, while it may be a legitimate scientific inquiry, is not a kind of epistemology, and, therefore, that the question whether it is a better kind of epistemology cannot arise. In reply, however, it might be said that there was a sense in which Quine's epistemology and traditional epistemology could be viewed as sharing a common subject matter, namely this: they both concern beliefs or "representations." The only difference is that the former investigates their causal histories and connections whereas the latter is concerned with their evidential or justificatory properties and relations. This difference, if Quine is right, leads to another (so continues the reply): the former is a feasible inquiry, the latter is not.

I now want to take my argument a step further: I shall argue that the concept of belief is itself an essentially normative one, and in consequence that if normativity is wholly excluded from naturalized epistemology it cannot even be thought of as being about beliefs. That is, if naturalized epistemology is to be a science of beliefs properly so called, it must presuppose a normative concept of belief.

Briefly, the argument is this. In order to implement Quine's program of naturalized epistemology, we shall need to identify, and individuate, the input and output of cognizers. The input, for Quine, consists of physical events ("the stimulation of sensory receptors") and the output is said to be a "theory" or "picture of the world" – that is, a set of "representations" of the cognizer's environment. Let us focus on the output. In order to study the sensory input–cognitive output relations for the given cognizer, therefore, we must find out what "representations" he has formed as a result of the particular stimulations that have been applied to his sensory transducers. Setting aside the jargon, what we need to be able to do is to attribute *beliefs*, and other contentful intentional states, to the cognizer. But belief attribution ultimately requires a "radical interpretation" of the cognizer, of his speech and intentional states; that is, we must construct an "interpretive theory" that simultaneously assigns meanings to his utterances and attributes to him beliefs and other propositional attitudes.²²

Even a cursory consideration indicates that such an interpretation cannot begin – we cannot get a foothold in our subject's realm of meanings and intentional states – unless we assume his total system of beliefs and other propositional attitudes to be largely and essentially rational and coherent. As Davidson has emphasized, a given belief has the content it has in part because of its location in a network of other beliefs and propositional attitudes; and what at bottom grounds this network is the evidential relation, a relation that regulates what is reasonable to believe given other beliefs one holds. That is, unless our cognizer is a "rational being," a being whose cognitive "output" is regulated and constrained by norms of rationality – typically, these norms holistically constrain his propositional attitudes in virtue of their contents – we cannot intelligibly interpret his "output" as consisting of beliefs. Conversely, if we are unable to interpret our subject's meanings and propositional attitudes in a way that satisfies a minimal standard of rationality, there is little reason to regard him as a "cognizer," a being that forms representations and constructs theories. This means that there is a sense of "rational" in which the expression "rational belief" is redundant; every belief must be rational in certain minimal ways. It is not important for the purposes of the present argument what these minimal standards of rationality are; the only point that matters

is that unless the output of our cognizer is subject to evaluation in accordance with norms of rationality, that output cannot be considered as consisting of beliefs and hence cannot be the object of an epistemological inquiry, whether plain or naturalized.

We can separate the core of these considerations from controversial issues involving the so-called "principle of charity," minimal rationality, and other matters in the theory of radical interpretation. What is crucial is this: for the interpretation and attribution of beliefs to be possible, not only must we assume the overall rationality of cognizers, but also we must continually evaluate and re-evaluate the putative beliefs of a cognizer in their evidential relationship to one another and other propositional attitudes. It is not merely that belief attribution requires the umbrella assumption about the overall rationality of cognizers. Rather, the point is that *belief attribution requires belief evaluation*, in accordance with normative standards of evidence and justification. If this is *correct*, rationality in its broad and fundamental sense is not an optional property of beliefs, a virtue that some beliefs may enjoy and others lack; it is a precondition of the attribution and individuation of belief – that is, a property without which the concept of belief would be unintelligible and pointless.

Two objections might be raised to counter these considerations. First, one might argue that at best they show only that the normativity of belief is an epistemological assumption – that we need to assume the rationality and coherence of belief systems when we are trying to *find out* what beliefs to attribute to a cognizer. It does not follow from this epistemological point, the objection continues, that the concept of belief is itself normative.²³ In replying to this objection, we can bypass the entire issue of whether the rationality assumption concerns only the epistemology of belief attribution. Even if this premise (which I think is incorrect) is granted, the point has already been made. For it is an essential part of the business of naturalized epistemology, as a theory of how beliefs are formed as a result of sensory stimulation, to *find out* what particular beliefs the given cognizers have formed. But this is precisely what cannot be done, if our considerations show anything at all, unless the would-be naturalized epistemologist continually evaluates the putative beliefs of his subjects in regard to their rationality and coherence, subject to the overall constraint of the assumption that the

cognizers are largely rational. The naturalized epistemologist cannot dispense with normative concepts or disengage himself from valuational activities.

Second, it might be thought that we could simply avoid these considerations stemming from belief attribution by refusing to think of cognitive output as consisting of “beliefs,” namely as states having propositional contents. The “representations” Quine speaks of should be taken as appropriate neural states, and this means that all we need is to be able to discern neural states of organisms. This requires only neurophysiology and the like, not the normative theory of rational belief. My reply takes the form of a dilemma: either the “appropriate” neural states are identified by seeing how they correlate with beliefs,²⁴ in which case we still need to contend with the problem of radical interpretation, or beliefs are entirely bypassed. In the latter case, belief, along with justification, drops out of Quinean epistemology, and it is unclear in what sense we are left with an inquiry that has anything to do with knowledge.²⁵

The “Psychologistic” Approach to Epistemology

Many philosophers now working in theory of knowledge have stressed the importance of systematic psychology to philosophical epistemology. Reasons proffered for this are various, and so are the conceptions of the proper relationship between psychology and epistemology.²⁶ But they are virtually unanimous in their rejection of what they take to be the epistemological tradition of Descartes and its modern embodiments in philosophers like Russell, C. I. Lewis, Roderick Chisholm, and A. J. Ayer; and they are united in their endorsement of the naturalistic approach of Quine we have been considering. Traditional epistemology is often condemned as “aprioristic,” and as having lost sight of human knowledge as a product of natural causal processes and its function in the survival of the organism and the species. Sometimes, the adherents of the traditional approach are taken to task for their implicit anti-scientific bias or indifference to the new developments in psychology and related disciplines. Their own approach in contrast is hailed as “naturalistic” and “scientific,” better attuned to significant advances in the relevant scientific fields such as “cognitive science” and “neuroscience,” promis-

ing philosophical returns far richer than what the aprioristic method of traditional epistemology has been able to deliver. We shall here briefly consider how this new naturalism in epistemology is to be understood in relation to the classic epistemological program and Quine’s naturalized epistemology.

Let us see how one articulate proponent of the new approach explains the distinctiveness of his position vis-à-vis that of the traditional epistemologists. According to Philip Kitcher, the approach he rejects is characterized by an “apsychologistic” attitude that takes the difference between knowledge and true belief – that is, justification – to consist in “ways which are independent of the causal antecedents of a subject’s states.”²⁷ Kitcher writes:²⁸

we can present the heart of [the apychologistic approach] by considering the way in which it would tackle the question of whether a person’s true belief that *p* counts as knowledge that *p*. The idea would be to disregard the psychological life of the subject, looking just at the various propositions she believes. If *p* is “connected in the right way” to other propositions which are believed, then we count the subject as knowing that *p*. Of course, apychologistic epistemology will have to supply a criterion for propositions to be “connected in the right way” . . . but proponents of this view of knowledge will emphasize that the criterion is to be given in *logical* terms. We are concerned with logical relations among propositions, not with psychological relations among mental states.

On the other hand, the psychologistic approach considers the crucial difference between knowledge and true belief – that is, epistemic justification – to turn on “the factors which produced the belief,” focusing on “processes which produce belief, processes which will always contain, at their latter end, psychological events.”²⁹

It is not entirely clear from this characterization whether a psychologistic theory of justification is to be *prohibited* from making *any* reference to logical relations among belief contents (it is difficult to believe how a theory of justification respecting such a blanket prohibition could succeed); nor is it clear whether, conversely, an apychologistic theory will be permitted to refer at all to beliefs qua psychological states, or exactly what it is for a theory to do so. But such points of detail are unimportant here; it is clear enough, for

example, that Goldman's proposal to explicate justified belief as belief generated by a reliable belief-forming process³⁰ nicely fits Kitcher's characterization of the psychologistic approach. This account, one form of the so-called "reliability theory" of justification, probably was what Kitcher had in mind when he was formulating his general characterization of epistemological naturalism. However, another influential form of the reliability theory does not qualify under Kitcher's characterization. This is Armstrong's proposal to explain the difference between knowledge and true belief, at least for noninferential knowledge, in terms of "a law-like connection between the state of affairs [of a subject's believing that *p*] and the state of affairs that makes '*p*' true such that, given the state of affairs [of the subject's believing that *p*], it must be the case that *p*."³¹ There is here no reference to the causal antecedents of beliefs, something that Kitcher requires of psychologicistic theories.

Perhaps, Kitcher's preliminary characterization needs to be broadened and sharpened. However, a salient characteristic of the naturalistic approach has already emerged, which we can put as follows: justification is to be characterized in terms of causal or nomological connections involving beliefs as psychological states or processes, and not in terms of the logical properties or relations pertaining to the contents of these beliefs.³²

If we understand current epistemological naturalism in this way, how closely is it related to Quine's conception of naturalized epistemology? The answer, I think, is obvious: not very closely at all. In fact, it seems a good deal closer to the Cartesian tradition than to Quine. For, as we saw, the difference that matters between Quine's epistemological program and the traditional program is the former's total renouncement of the latter's normativity, its rejection of epistemology as a normative inquiry. The talk of "replacing" epistemology with psychology is irrelevant and at best misleading, though it could give us a momentary relief from a sense of deprivation. When one abandons justification and other valuational concepts, one abandons the entire framework of normative epistemology. What remains is a descriptive empirical theory of human cognition which, if Quine has his way, will be entirely devoid of the notion of justification or any other evaluative concept.

As I take it, this is not what most advocates of epistemological naturalism are aiming at. By and large they are not Quinean eliminativists in regard

to justification, and justification in its full-fledged normative sense continues to play a central role in their epistemological reflections. Where they differ from their nonnaturalist adversaries is the specific way in which criteria of justification are to be formulated. Naturalists and nonnaturalists ("apsychologists") can agree that these criteria must be stated in descriptive terms – that is, without the use of epistemic or any other kind of normative terms. According to Kitcher, an psychologicistic theory of justification would state them primarily in terms of logical properties and relations holding for propositional contents of beliefs, whereas the psychologistic approach advocates the exclusive use of causal properties and relations holding for beliefs as events or states. Many traditional epistemologists may prefer criteria that confer upon a cognizer a position of special privilege and responsibility with regard to the epistemic status of his beliefs, whereas most self-avowed naturalists prefer "objective" or "externalist" criteria with no such special privileges for the cognizer. But these differences are among those that arise within the familiar normative framework, and are consistent with the exclusion of normative terms in the statement of the criteria of justification.

Normative ethics can serve as a useful model here. To claim that basic ethical terms, like "good" and "right," are definable on the basis of descriptive or naturalistic terms is one thing; to insist that it is the business of normative ethics to provide conditions or criteria for "good" and "right" in descriptive or naturalistic terms is another. One may properly reject the former, the so-called "ethical naturalism," as many moral philosophers have done, and hold the latter; there is no obvious inconsistency here. G. E. Moore is a philosopher who did just that. As is well known, he was a powerful critic of ethical naturalism, holding that goodness is a "simple" and "nonnatural" property. At the same time, he held that a thing's being good "follows" from its possessing certain naturalistic properties. He wrote:³³

I should never have thought of suggesting that goodness was "non-natural," unless I had supposed that it was "derivative" in the sense that, whenever a thing is good (in the sense in question) its goodness... "depends on the presence of certain non-ethical characteristics" possessed by the thing in question: I have always supposed that it did so "depend," in the sense that, if a thing is good (in my sense), then that it is so

follows from the fact that it possesses certain natural intrinsic properties...

It makes sense to think of these “natural intrinsic properties” from which a thing’s being good is thought to follow as constituting naturalistic criteria of goodness, or at least pointing to the existence of such criteria. One can reject ethical naturalism, the doctrine that ethical concepts are definitionally eliminable in favor of naturalistic terms, and at the same time hold that ethical properties, or the ascription of ethical terms, must be governed by naturalistic criteria. It is clear, then, that we are here using “naturalism” ambiguously in “epistemological naturalism” and “ethical naturalism.” In our present usage, epistemological naturalism does not include (nor does it necessarily exclude) the claim that epistemic terms are definitionally reducible to naturalistic terms. (Quine’s naturalism is eliminative, though it is not a definitional eliminativism.)

If, therefore, we locate the split between Quine and traditional epistemology at the descriptive vs. normative divide, then currently influential naturalism in epistemology is not likely to fall on Quine’s side. On this descriptive vs. normative issue, one can side with Quine in one of two ways: first, one rejects, with Quine, the entire justification-based epistemological program; or second, like ethical naturalists but unlike Quine, one believes that epistemic concepts are naturalistically definable. I doubt that very many epistemological naturalists will embrace either of these alternatives.³⁴

Epistemic Supervenience – Or Why Normative Epistemology Is Possible

But why should we think that there *must be* naturalistic criteria of justified belief and other terms of epistemic appraisal? If we take the discovery and systematization of such criteria to be the central task of normative epistemology, is there any reason to think that this task can be fruitfully pursued, that normative epistemology is a possible field of inquiry? Quine’s point is that it is not. We have already noted the limitation of Quine’s negative arguments in “Epistemology Naturalized,” but is there a positive reason for thinking that normative epistemology is a viable program? One could consider a similar question about the possibility of normative ethics.

I think there is a short and plausible initial answer, although a detailed defense of it would involve complex general issues about norms and values. The short answer is this: we believe in the supervenience of epistemic properties on naturalistic ones, and more generally, in the supervenience of all valuational and normative properties on naturalistic conditions. This comes out in various ways. We think, with R. M. Hare,³⁵ that if two persons or acts coincide in all descriptive or naturalistic details, they cannot differ in respect of being good or right, or any other valuational aspects. We also think that if something is “good” – a “good car,” “good drop shot,” “good argument” – then that must be so “in virtue of” its being a “certain way,” that is, its having certain “factual properties.” Being a good car, say, cannot be a brute and ultimate fact: a car is good *because* it has a certain contextually indicated set of properties having to do with performance, reliability, comfort, styling, economy, etc. The same goes for justified belief: if a belief is justified, that must be so *because* it has a certain factual, non-epistemic properties, such as perhaps that it is “indubitable,” that it is seen to be entailed by another belief that is independently justified, that it is appropriately caused by perceptual experience, or whatever. That it is a justified belief cannot be a brute fundamental fact unrelated to the kind of belief it is. There must be a *reason* for it, and this reason must be grounded in the factual descriptive properties of that particular belief. Something like this, I think, is what we believe.

Two important themes underlie these convictions: first, values, though perhaps not reducible to facts, must be “consistent” with them in that objects that are indiscernible in regard to fact must be indiscernible in regard to value; second, there must be nonvaluational “reasons” or “grounds” for the attribution of values, and these “reasons” or “grounds” must be *generalizable* – that is, they are covered by *rules* or *norms*. These two ideas correspond to “weak supervenience” and “strong supervenience” that I have discussed elsewhere.³⁶ Belief in the supervenience of value upon fact, arguably, is fundamental to the very concepts of value and valuation.³⁷ Any valuational concept, to be significant, must be governed by a set of criteria, and these criteria must ultimately rest on factual characteristics and relationships of objects and events being evaluated. There is something deeply incoherent about the idea of an infinitely descending series of valuational concepts,

each depending on the one below it as its criterion of application.³⁸

It seems to me, therefore, that epistemological supervenience is what underlies our belief in the possibility of normative epistemology, and that we do not need new inspirations from the sciences to acknowledge the existence of naturalistic criteria for epistemic and other valuational concepts. The case of normative ethics is entirely parallel: belief in the possibility of normative ethics is rooted in the belief that moral properties and relations are supervenient upon nonmoral ones. Unless we are prepared to disown normative ethics as a viable philosophical inquiry, we had better recognize normative epistemology as one, too.³⁹ We should

note, too, that epistemology is likely to parallel normative ethics in regard to the degree to which scientific results are relevant or useful to its development.⁴⁰ Saying this of course leaves large room for disagreement concerning how relevant and useful, if at all, empirical psychology of human motivation and action can be to the development and confirmation of normative ethical theories.⁴¹ In any event, once the normativity of epistemology is clearly taken note of, it is no surprise that epistemology and normative ethics share the same metaphilosophical fate. Naturalized epistemology makes no more, and no less, sense than naturalized normative ethics.⁴²

Notes

- 1 In making these remarks I am only repeating the familiar textbook history of philosophy; however, what *our* textbooks say about the history of a philosophical concept has much to do with *our* understanding of that concept.
- 2 Goldman 1979 explicitly states this requirement as a desideratum of his own analysis of justified belief. Chisholm's 1977 definition of "being evident" does not satisfy this requirement as it rests ultimately on an unanalyzed epistemic concept of one belief being *more reasonable than* another. What does the real "criteriological" work for Chisholm is his "principles of evidence." See especially (A) on p. 73 of his 1977, which can usefully be regarded as an attempt to provide nonnormative, descriptive conditions for certain types of justified beliefs.
- 3 The basic idea of this stronger requirement seems implicit in Firth's notion of "warrant-increasing property" in his 1964. It seems that Alston 1976 has something similar in mind when he says, "like any evaluative property, epistemic justification is a supervenient property, the application of which is based on more fundamental properties" (at this point Alston refers to Firth's paper cited above) (the quoted remark occurs on p. 170). Although Alston doesn't further explain what he means by "more fundamental properties," the context makes it plausible to suppose that he has in mind non-normative, descriptive properties. See section 7 below for further discussion.
- 4 See Chisholm 1977, p. 14. Here Chisholm refers to a "person's responsibility or duty *qua* intellectual being."
- 5 This term was used by Ayer 1956 to characterize the difference between lucky guessing and knowing, p. 33.
- 6 Notably by Chisholm in 1977, 1st edn, ch. 4.
- 7 See Carnap, 1936. We should also note the presence of a strong coherentist streak among some positivists; see, e.g., Hempel 1935.
- 8 In Quine 1969; see this vol., ch. 23. Also see his 1960; 1973; 1970; and especially 1975. See Schmitt's excellent bibliography on naturalistic epistemology in Kornblith 1985.
- 9 Or conformational relations, given the Positivists' verificationist theory of meaning.
- 10 I know of no serious defense of it since Ayer's 1940.
- 11 See Kornblith 1985a, pp. 19–20.
- 12 *Ibid.*, p. 21.
- 13 To use an expression of Rorty's 1979, p. 11.
- 14 Sober 1978 makes a similar point: "And on the question of whether the failure of a foundationalist programme shows that questions of justification cannot be answered, it is worth noting that Quine's advice 'Since Carnap's foundationalism failed, why not settle for psychology' carries weight only to the degree that Carnapian epistemology exhausts the possibilities of epistemology."
- 15 See Chisholm 1977, ch. 4.
- 16 "If we are seeking only the causal mechanism of our knowledge of the external world, and not a justification of that knowledge in terms prior to science . . ." Quine 1970, p. 2.
- 17 *Ibid.*, p. 75. Emphasis added.
- 18 *Ibid.*, p. 78. Emphasis added.
- 19 *Ibid.*, p. 83. Emphasis added.
- 20 But aren't there those who advocate a "causal theory" of evidence or justification? I want to make two brief points about this. First, the nomological or causal input-output relations are not in themselves evidential relations, whether these latter are understood causally or otherwise. Second, a causal theory of evidence attempts to state *criteria* for "e is evidence for h" in causal terms; even if this is success-

- ful, it does not necessarily give us a causal "definition" or "reduction" of the concept of evidence. For more details see section 6 below.
- 21 I am not saying that Quine is under any illusion on this point. My remarks are directed rather at those who endorse Quine without, it seems, a clear appreciation of what is involved.
 - 22 Here I am drawing chiefly on Davidson's writings on radical interpretation. See Essays 9, 10, and 11 in his 1984. See also Lewis 1974.
 - 23 Robert Audi suggested this as a possible objection.
 - 24 For some considerations tending to show that these correlations cannot be lawlike, see my 1985.
 - 25 For a more sympathetic account of Quine than mine, see Kornblith's introductory essay in his 1985.
 - 26 See, for more details, Goldman 1986.
 - 27 Kitcher 1983, p. 14.
 - 28 Ibid.
 - 29 Ibid., p. 13. I should note that Kitcher considers the psychologistic approach to be an aberration of the twentieth-century epistemology, as represented by philosophers like Russell, Moore, C. I. Lewis, and Chisholm, rather than an historical characteristic of the Cartesian tradition. Kornblith 1982 gives an analogous characterization of the two approaches to justification; he associates "justification-conferring processes" with the psychologistic approach and "epistemic rules" with the apsychologistic approach.
 - 30 See Goldman 1979.
 - 31 Armstrong 1973, p. 166.
 - 32 The aptness of this characterization of the "apsychologistic" approach for philosophers like Russell, Chisholm, Lehrer, Pollock, etc. can be debated. Also, there is the issue of "internalism" vs. "externalism" concerning justification, which I believe must be distinguished from the psychologistic vs. apsychologistic division.
 - 33 Moore, 1942, p. 588.
 - 34 Rorty's claim, which plays a prominent role in his arguments against traditional epistemology, that Locke and other modern epistemologists conflated the normative concept of justification with causal-mechanical concepts is itself based, I believe, on a conflation of just the kind I am describing here. See Rorty, 1979, pp. 139ff. Again, the critical conflation consists in not seeing that the view, which I believe is correct, that epistemic justification, like any other normative concept, must have factual, naturalistic criteria, is entirely consistent with the rejection of the doctrine, which I think is incorrect, that justification *is*, or is *reducible* to, a naturalistic-nonnormative concept.
 - 35 Hare 1952, p. 145.
 - 36 See Kim 1984.
 - 37 Sosa, too, considers epistemological supervenience as a special case of the supervenience of valuational properties on naturalistic conditions in his 1980, especially p. 551. See also Van Cleve's instructive discussion in his 1985, especially, pp. 97-9.
 - 38 Perhaps one could avoid this kind of criteriological regress by embracing directly apprehended valuational properties (as in ethical intuitionism) on the basis of which criteria for other valuational properties could be formulated. The denial of the supervenience of valuational concepts on factual characteristics, however, would sever the essential connection between value and fact on which, it seems, the whole point of our valuational activities depends. In the absence of such supervenience, the very notion of valuation would lose its significance and relevance. The elaboration of these points, however, would have to wait for another occasion; but see Van Cleve's paper cited in the preceding note for more details.
 - 39 Quine will not disagree with this: he will "naturalize" them both. For his views on values see 1978. For a discussion of the relationship between epistemic and ethical concepts see Firth 1978.
 - 40 For discussions of this and related issues see Goldman 1986.
 - 41 For a detailed development of a normative ethical theory that exemplifies the view that it is crucially relevant, see Brandt 1979.
 - 42 An earlier version of this paper was read at a meeting of the Korean Society for Analytic Philosophy in 1984 in Seoul. An expanded version was presented at a symposium at the Western Division meetings of the American Philosophical Association in April, 1985, and at the epistemology conference at Brown University in honor of Roderick Chisholm in 1986. I am grateful to Richard Foley and Robert Audi who presented helpful comments at the APA session and the Chisholm Conference respectively. I am also indebted to Terence Horgan and Robert Meyers for helpful comments and suggestions.

References

- Alston, William, 1976. "Two Types of Foundationalism," *Journal of Philosophy* 73, pp. 165-85.
- Armstrong, David M., 1973. *Truth, Belief and Knowledge* (London: Cambridge University Press).
- Ayer, A. J., 1940. *The Foundations of Empirical Knowledge* (London: Macmillan).
- , 1956. *The Problem of Knowledge* (London: Penguin Books).

- Brandt, Richard, 1979. *A Theory of the Good and the Right* (Oxford: Clarendon Press).
- Carnap, Rudolf, 1936. "Testability and Meaning," *Philosophy of Science* 3.
- Chisholm, Roderick M., 1977. *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall).
- Davidson, Donald, 1984. *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press).
- Firth, Roderick, 1964. "Coherence, Certainty, and Epistemic Priority," *Journal of Philosophy* 61, pp. 545–57.
- , 1978. "Are Epistemic Concepts Reducible to Ethical Concepts?" in A. I. Goldman and J. Kim (eds), *Values and Morals* (Dordrecht: Reidel).
- Goldman, Alvin I., 1979. "What Is Justified Belief?" this vol., ch. 27.
- , 1986. *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- Hare, R. M., 1952. *The Language of Morals* (London: Oxford University Press).
- Hempel, Carl G., 1935. "Some Remarks on 'Facts' and Propositions," *Analysis* 2, pp. 93–6.
- Kim, Jaegwon, 1984. "Concepts of Supervenience," *Philosophy and Phenomenological Research* 45, pp. 153–76.
- , 1985. "Psychophysical Laws," in Ernest LePore and Brian McLaughlin (eds), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell).
- Kitcher, Philip, 1983. *The Nature of Mathematical Knowledge* (New York: Oxford University Press).
- Kornblith, Hilary, 1985. "What Is Naturalistic Epistemology?" in Hilary Kornblith (ed.), *Naturalizing Epistemology* (Cambridge, MA: MIT Press).
- , 1985a. *Naturalizing Epistemology* (Cambridge, MA: MIT Press).
- Lewis, David, 1974. "Radical Interpretation," *Synthese* 27, pp. 331–44.
- Moore, G. E., 1942. "A Reply to My Critics," in P. A. Schilpp (ed.), *The Philosophy of G. E. Moore* (Chicago & Evanston: Open Court).
- Quine, W. V., 1960. *Word and Object* (Cambridge, MA: MIT Press).
- , 1969. *Ontological Relativity and Other Essays* (New York: Columbia University Press).
- , 1970. "Grades of Theoreticity," in L. Foster and J. W. Swanson (eds), *Experience and Theory* (Amherst, MA: University of Massachusetts Press).
- , 1973. *The Roots of Reference* (La Salle, IL: Open Court).
- , 1975. "The Nature of Natural Knowledge," in Samuel Guttenplan (ed), *Mind and Language* (Oxford: Clarendon Press).
- , 1978. "The Nature of Moral Values," in Alvin I. Goldman and Jaegwon Kim (eds), *Values and Morals* (Dordrecht: Reidel).
- Rorty, Richard, 1979. *Philosophy and the Mirror Of Nature* (Princeton, NJ: Princeton University Press).
- Schmitt, Frederick, 1985. Bibliography, in Hilary Kornblith (ed.), *Naturalizing Epistemology*. Cambridge, MA: MIT Press.
- Sober, Elliott, 1978. "Psychologism," *Journal of Theory of Social Behavior* 8, pp. 165–91.
- Sosa, Ernest, 1980. "The Foundations of Foundationalism," *Nous* 14, pp. 547–64.
- Van Cleve, James, 1985. "Epistemic Supervenience and the Circle of Belief," *The Monist* 68, pp. 90–104.

Why Reason Can't Be Naturalized

Hilary Putnam

In this chapter I shall examine attempts to naturalize the fundamental notions of the theory of knowledge, for example the notion of a belief's being *justified* or *rationaly acceptable*.

While the two sorts of attempts are alike in that they both seek to reduce 'intentional' or mentalistic notions to materialistic ones, and thus are both manifestations of what Peter Strawson (1979) has described as a permanent tension in philosophy, in other ways they are quite different. The materialist metaphysician often uses such traditional metaphysical notions as *causal power*, and *nature* quite uncritically. (I have even read papers in which one finds the locution 'realist truth', as if everyone understood this notion except a few fuzzy anti-realists.) The 'physicalist' generally doesn't seek to *clarify* these traditional metaphysical notions, but just to show that *science* is progressively verifying the *true* metaphysics. That is why it seems just to describe *his* enterprise as 'natural metaphysics', in strict analogy to the 'natural theology' of the eighteenth and nineteenth centuries. Those who raise the slogan 'epistemology naturalized', on the other hand, generally *disparage* the traditional enterprises of epistemology. In this respect, moreover, they do not differ from philosophers of a less reductionist kind; the criticism they voice of traditional epistemology – that it was in the grip of a 'quest for certainty', that it was unrealistic in seeking a 'foundation' for knowledge as a whole, that the 'foundation' it claimed to provide was by

no means indubtable in the way it claimed, that the whole 'Cartesian enterprise' was a mistake, etc., – are precisely the criticisms one hears from philosophers of all countries and types. Hegel already denounced the idea of an 'Archimedean point' from which epistemology could judge all of our scientific, legal, moral, religious, etc. beliefs (and set up standards for all of the special subjects). It is true that Russell and Moore ignored these strictures of Hegel (as they ignored Kant), and revived 'foundationalist epistemology'; but today that enterprise has few defenders. The fact that the naturalized epistemologist is trying to reconstruct what he can of an enterprise that few philosophers of any persuasion regard as unflawed is perhaps the explanation of the fact that the naturalistic tendency in epistemology expresses itself in so many incompatible and mutually divergent ways, while the naturalistic tendency in metaphysics appears to be, and regards itself as, a unified movement.

Evolutionary Epistemology

The simplest approach to the problem of giving a naturalistic account of reason is to appeal to Darwinian evolution. In its crudest form, the story is familiar: reason is a capacity we have for discovering truths. Such a capacity has survival value; it evolved in just the way that any of our physical organs or capacities evolved. A belief is rational if it is arrived at by the exercise of this capacity.

This approach assumes, at bottom, a metaphysically 'realist' notion of truth: truth as 'correspondence to the facts' or something of that kind.

Originally published in H. Putnam, *Realism and Reason*, vol. 3 of *Philosophical Papers* (Cambridge, MA: Harvard University Press), pp. 229–47; reprinted by permission of the publisher, copyright © by the President and Fellows of Harvard College.

And this notion is incoherent. We don't have notions of the 'existence' of things or of the 'truth' of statements that are independent of the versions we construct and of the procedures and practices that give sense to talk of 'existence' and 'truth' within those versions. Do *fields* 'exist' as physically real things? Yes, fields really exist: relative to one scheme for describing and explaining physical phenomena; relative to another there are particles, plus 'virtual' particles, plus 'ghost' particles, plus... Is it true that *brown* objects exist? Yes, relative to a common-sense version of the world: although one cannot give a necessary and sufficient condition for an object to be brown,¹ (one that applies to all objects, under all conditions) in the form of a finite closed formula in the language of physics. Do *dispositions* exist? Yes, in our ordinary way of talking (although disposition talk is just as recalcitrant to translation into physicalistic language as counterfactual talk, and for similar reasons). We have many irreducibly different but legitimate ways of talking, and true 'existence' statements in all of them.

To postulate a set of 'ultimate' objects, the furniture of the world, or what you will, whose 'existence' is *absolute*, not relative to our discourse at all, and a notion of truth as 'correspondence' to these ultimate objects is simply to revive the whole failed enterprise of traditional metaphysics.

Truth, in the only sense in which we have a vital and working notion of it, is rational acceptability (or, rather, rational acceptability under sufficiently good epistemic conditions; and which conditions are epistemically better or worse is relative to the type of discourse in just the way rational acceptability itself is). But to substitute this characterization of truth into the formula 'reason is a capacity for discovering truths' is to see the emptiness of that formula at once: 'reason is a capacity for discovering what is (or would be) rationally acceptable' is *not* the most informative statement a philosopher might utter. The evolutionary epistemologist must either presuppose a 'realist' (i.e., a metaphysical) notion of truth or see his formula collapse into vacuity.

Roderick Firth² has argued that, in fact, it collapses into a kind of epistemic vacuity on *any* theory of rational acceptability (or truth). For, he points out, whatever we take the correct epistemology (or the correct theory of truth) to be, we have no way of *identifying* truths except to posit that the statements that are currently rationally acceptable (by our lights) are true. Even if these beliefs are

false, even if our rational beliefs contribute to our survival for some reason *other* than truth, the way 'truths' are identified *guarantees* that reason will seem to be a 'capacity for discovering truths'. This characterization of reason has thus no real empirical content.

The evolutionary epistemologist could, I suppose, try using some notion *other* than the notion of 'discovering truths'. For example, he might try saying that 'reason is a capacity for arriving at beliefs which *promote our survival*' (or our 'inclusive genetic fitness'). But this would be a loser! Science itself, and the methodology which we have developed since the seventeenth century for constructing and evaluating theories, has *mixed* effects on inclusive genetic fitness and all too uncertain effects on survival. If the human race perishes in a nuclear war, it may well be (although there will be no one alive to say it) that scientific beliefs did *not*, in a sufficiently long time scale, promote 'survival'. Yet that will not have been because the scientific theories were not rationally acceptable, but because our *use* of them was irrational. In fact, if rationality were measured by survival value, then the proto-beliefs of the cockroach, who has been around for tens of millions of years longer than we, would have a far higher claim to rationality than the sum total of human knowledge. But such a measure would be cock-eyed; there is no contradiction in imagining a world in which people have utterly irrational beliefs which for some reason enable them to survive, or a world in which the most rational beliefs quickly lead to extinction.

If the notion of 'truth' in the characterization of rationality as a 'capacity for discovering truths' is problematic, so, almost equally, is the notion of a 'capacity'. In one sense of the term, *learning* is a 'capacity' (even, a 'capacity for discovering truths'), and *all* our beliefs are the product of *that* capacity. Yet, for better or worse, not all our beliefs are rational.

The problem here is that there are no sharp lines in the brain between one 'capacity' and another (Chomskians to the contrary). Even seeing includes not just the visual organs, the eyes, but the whole brain; and what is true of seeing is certainly true of *thinking* and *inferring*. We draw lines between one 'capacity' and another (or build them into the various versions we construct); but a sharp line at one level does not usually correspond to a sharp line at a lower level. The table at which I write, for example, is a natural unit at the level of

everyday talk; I am aware that the little particle of food sticking to its surface (I must do something about that!) is not a 'part' of the table; but at the physicist's level, the decision to consider that bit of food to be outside the boundary of the table is not natural at all. Similarly, 'believing' and 'seeing' are quite different at the level of ordinary language psychology (and usefully so); but the corresponding brain-processes interpenetrate in complex ways which can only be separated by looking outside the brain, at the environment and at the output behavior *as structured by our interests and salencies*. 'Reason is a capacity' is what Wittgenstein called a 'grammatical remark'; by which he meant (I think) not an analytic truth, but simply the sort of remark that philosophers often *take* to be informative when in fact it tells us nothing useful.

None of this is intended to deny the obvious scientific facts: that we would not be able to reason if we did not have brains, and that those *brains* are the product of evolution by natural selection. What is wrong with evolutionary epistemology is not that the scientific facts are wrong, but that they don't answer any of the philosophical questions.

The Reliability Theory of Rationality

A more sophisticated recent approach to these matters, proposed by Professor Alvin Goldman (1979), runs as follows: let us call a *method* (as opposed to a single belief) *reliable* if the method leads to a high frequency (say, 95 per cent) of *true* beliefs in a long-run series of representative applications (or *would* lead to such a high truth-frequency in such a series of applications). Then (the proposal goes) we can define a *rational* belief to be one which is *arrived at by using a reliable method*.

This proposal does not avoid the first objection we raised against evolutionary epistemology: it too presupposes a metaphysical notion of truth. Forgetting that rational acceptability does the lion's share of the work in fixing the notion of 'truth', the reliability theorist only pretends to be giving an analysis of rationality in terms that do not presuppose it. The second objection we raised against evolutionary epistemology, namely that the notion of a 'capacity' is hopelessly vague and general, is met, however, by replacing that notion with the notion of an arbitrary method for generating true or false statements, and then restricting the class to

those methods (in this sense) whose reliability (as defined) is high. 'Learning' may be a method for generating statements, but its *reliability* is not high enough for every statement we 'learn' to count as rationally acceptable, on this theory. Finally, *no* hypothesis is made as to whether the reliable methods we employ are the result of biological evolution, cultural evolution, or what: this is regarded as no part of the theory of what rationality *is*, in this account.

This account is vulnerable to many counterexamples, however. *One* is the following: suppose that Tibetan Buddhism is, in fact, *true*, and that the Dalai Lama is, in fact, *infallible* on matter of faith and morals. Anyone who believes in the Dalai Lama, and who invariably believes any statement the Dalai Lama makes on a matter of faith or morals, follows a method which is 100 per cent reliable; thus, if the reliability theory of rationality were correct, such a person's beliefs on faith and morals would all be rational *even if his argument for his belief that the Dalai Lama is never wrong is 'the Dalai Lama says so'*.

Cultural Relativism

I have already said that, in my view, truth and rational acceptability – a claim's being right and someone's being in a position to make it – are relative to the sort of language we are using and the sort of context we are in. 'That weighs one pound' may be true in a butcher shop, but the same sentence would be understood very differently (as demanding four decimal places of precision, perhaps) if the same object were being weighed in a laboratory. This does not mean that a claim is right *whenever* those who employ the language in question would accept it as right in its context, however. There are two points that must be *balanced*, both points that have been made by philosophers of many different kinds: (1) talk of what is 'right' and 'wrong' in any area only makes sense against the background of an *inherited tradition*; but (2) traditions themselves can be *criticized*. As Austin (1961) says, remarking on a special case of this, 'superstition and error and fantasy of all kinds do become incorporated in ordinary language and even sometimes stand up to the survival test (only, when they do, why should we not detect it?)'.

What I am saying is that the 'standards' accepted by a culture or a subculture, either explicitly or implicitly, cannot *define* what reason is,

even in context, because they *presuppose* reason (reasonableness) for their interpretation. On the one hand, there is no notion of reasonableness at all *without* cultures, practices, procedures; on the other hand, the cultures, practices, procedures we inherit are not an algorithm to be slavishly followed. As Mill said, commenting on his own inductive logic, there is no rule book which will not lead to terrible results 'if supposed to be conjoined with universal idiocy'. Reason is, in this sense, both immanent (not to be found outside of concrete language games and institutions) and transcendent (a regulative idea that we use to criticize the conduct of *all* activities and institutions).

Philosophers who lose sight of the immanence of reason, of the fact that reason is always relative to context and institution, become lost in characteristic philosophical fantasies. 'The ideal language', 'inductive logic', 'the empiricist criterion of significance' – these are the fantasies of the positivist, who would replace the vast complexity of human reason with a kind of intellectual Walden II. 'The absolute idea': this is the fantasy of Hegel, who, without ignoring that complexity, would have us (or, rather, 'spirit') reach an endstage at which we (it) could comprehend it all. Philosophers who lose sight of the transcendence of reason become cultural (or historical) relativists.

I want to talk about cultural relativism, because it is one of the most influential – perhaps the most influential – forms of naturalized epistemology extant, although not usually recognized as such.

The situation is complicated, because cultural relativists usually *deny* that they are cultural relativists. I shall count a philosopher as a cultural relativist for our purposes if I have not been able to find anyone who can explain to me why he *isn't* a cultural relativist. Thus I count Richard Rorty as a cultural relativist, because his explicit formulations are relativist ones (he identifies truth with right assertibility by the standards of one's cultural peers, for example), and because his entire attack on traditional philosophy is mounted on the basis that the nature of reason and representation are non-problems, because the only kind of truth it makes sense to seek is to convince one's cultural peers. Yet he himself *tells* us that relativism is self-refuting (Rorty, 1980). And I count Michel Foucault as a relativist because his insistence on the determination of beliefs by language is so overwhelming that it is an incoherence on his part not to apply his doctrine to his *own* language and

thought. Whether Heidegger ultimately escaped something very much like cultural, or rather historical, relativism is an interesting question.

Cultural relativists are not, in their own eyes, scientific or 'physicalistic'. They are likely to view materialism and scientism as just the hang-ups of one particular cultural epoch. If I count them as 'naturalized epistemologists' it is because their doctrine is, none the less, a product of the same deference to the claims of nature, the same desire for harmony with the world version of some science, as physicalism. The difference in style and tone is thus explained: the physicalist's paradigm of science is a *hard* science, *physics* (as the term 'physicalism' suggests); the cultural relativist's paradigm is a *soft* science: anthropology, or linguistics, or psychology, or history, as the case may be. That reason is whatever the norms of the local culture determine it to be is a naturalist view inspired by the *social* sciences, including history.

There is something which makes cultural relativism a far more dangerous cultural tendency than materialism. At bottom, there is a deep irrationalism to cultural relativism, a denial of the possibility of *thinking* (as opposed to making noises in counterpoint or in chorus). An aspect of this which is of special concern to philosophy is the suggestion, already mentioned, that the deep questions of philosophy are not deep at all. A corollary to this suggestion is that philosophy, as traditionally understood, is a *silly* enterprise. But the questions *are* deep, and it is the easy answers that are silly. Even seeing that relativism is inconsistent is, if the knowledge is taken seriously, seeing something important about a deep question. Philosophers *are* beginning to talk about the great issues again, and to feel that something can be *said* about them, even if there are no grand or ultimate solutions. There is an excitement in the air. And if I react to Professor Rorty's book (1979) with a certain sharpness, it is because one more 'deflationary' book, one more book telling us that the deep questions aren't deep and the whole enterprise was a mistake, is just what we *don't* need right now. Yet I am grateful to Rorty all the same, for his work has the merit of addressing profound questions head-on.

So, although we all know that cultural relativism is inconsistent (or say we do) I want to take the time to say again that it is inconsistent. I want to point out one reason that it is: not one of the quick, logic-chopping refutations (although every refutation of relativism teaches us something about reason) but a somewhat messy, somewhat 'intuitive', reason.

I shall develop my argument in analogy with a well-known argument against 'methodological solipsism'. The 'methodological solipsist' – one thinks of Carnap's *Logische Aufbau* or of Mach's *Analyse der Empfindungen* – holds that *all* our talk can be reduced to talk about experiences and logical constructions out of experiences. More precisely, he holds that everything he can conceive of is identical (in the ultimate logical analyses of his language) with one or another complex of his *own* experiences. What makes him a *methodological* solipsist as opposed to a real solipsist is that he kindly adds that *you*, dear reader, are the 'I' of this construction when *you* perform it: he says *everybody* is a (methodological) solipsist.

The trouble, which should be obvious, is that his two stances are ludicrously incompatible. His solipsist stance implies an enormous asymmetry between persons: my body is a construction out of my experiences, in the system, but *your* body isn't a construction out of *your* experiences. It's a construction out of *my* experiences. And your experiences – viewed from within the system – are a construction out of your bodily behavior, which, as just said, is a construction out of *my* experiences. My experiences are different from everyone else's (within the system) in that they are what *everything* is constructed from. But his transcendental stance is that it's all symmetrical: the 'you' he addresses his higher-order remark to cannot be the *empirical* 'you' of the system. But if it's really true that the 'you' of the system is the only 'you' he can *understand*, then the transcendental remark is *unintelligible*. Moral: don't be a methodological solipsist unless you are a *real* solipsist!

Consider now the position of the cultural relativist who says, 'When I say something is *true*, I mean that it is correct according to the norms of *my* culture.' If he adds, 'When a member of a different culture says that something is true, what he means (whether he knows it or not) is that it is in conformity with the norms of *his* culture', then he is in exactly the same plight as the methodological solipsist.

To spell this out, suppose R. R., a cultural relativist, says

When Karl says 'Schnee ist weiss', what Karl means (whether he knows it or not) is that snow is white *as determined* by the norms of Karl's culture

(which we take to be German culture).

Now the sentence 'Snow is white as determined by the norms of German culture' is itself one which R. R. has to *use*, not just mention, to say what Karl says. On his own account, what R. R. means by *this* sentence is

'Snow is white as determined by the norms of German culture' is true by the norms of R. R.'s culture

(which we take to be American culture).

Substituting this back into the first displayed utterance, (and changing to indirect quotation) yields:

When Karl says 'Schnee ist weiss', what he means (whether he knows it or not) is that it is true as determined by the norms of American culture that it is true as determined by the norms of German culture that snow is white.

In general, if R. R. understands *every* utterance *p* that *he* uses as meaning 'it is true by the norms of American culture that *p*', then he must understand his own hermeneutical utterances, the utterances he uses to interpret others, the same way, no matter how many qualifiers of the 'according to the norms of German culture' type or however many footnotes, glosses, commentaries on the cultural differences, or whatever, he accompanies them by. Other cultures become, so to speak, logical constructions out of the procedures and practices of American culture. If he now attempts to add 'the situation is reversed from the point of view of the *other* culture' he lands in the predicament the methodological solipsist found himself in: the transcendental claim of a *symmetrical* situation cannot be *understood* if the relativist doctrine is *right*. And to say, as relativists often do, that the other culture has 'incommensurable' concepts is no better. This is just the transcendental claim in a special jargon.

Stanley Cavell (1979, part IV) has written that skepticism about other minds can be a significant problem because we don't, in fact, always fully acknowledge the reality of others, their equal *validity* so to speak. One might say that the methodological solipsist is led to his transcendental observation that everyone is equally the 'I' of the construction by his praiseworthy desire to *acknowledge* others in this sense. But you *can't* acknowledge others in this sense, which involves recognizing that the situation *really* is symmetrical,

if you think they are really constructions out of *your* sense data. Nor can you acknowledge others in this sense if you think that the *only* notion of truth there is for *you* to understand is 'truth-as-determined-by-the-norms-of-this-culture'.

For simplicity, I have discussed relativism with respect to truth, but the same discussion applies to relativism about rational acceptability, justification, etc; indeed, a relativist is unlikely to be a relativist about one of these notions and not about the others.

Cultural Imperialism

Just as the methodological solipsist can become a *real* solipsist, the cultural relativist can become a cultural imperialist. He can say, 'Well then, truth – the only notion of truth I understand – is defined by the norms of *my* culture.' ('After all', he can add, 'which norms should I rely on? The norms of *somebody else's* culture?') Such a view is no longer relativist at all. It postulates an *objective* notion of truth, although one that is said to be a product of our culture, and to be defined by our culture's criteria (I assume the cultural imperialist is one of *us*). In this sense, just as consistent solipsism becomes indistinguishable from realism (as Wittgenstein said in the *Tractatus*), consistent cultural relativism also becomes indistinguishable from realism. But cultural imperialist realism is a special kind of realism.

It is realist in that it accepts an objective difference between what is true and what is merely thought to be true. (Whether it can consistently account for this difference is another question.)

It is not a *metaphysical* or transcendental realism, in that truth cannot go beyond right assertibility, as it does in metaphysical realism. But the notion of right assertibility is fixed by 'criteria', in a positivistic sense: something is rightly assertible only if the norms of the culture specify that it is; these norms are, as it were, an *operational definition* of right assertibility, in this view.

I don't know if any philosopher holds such a view, although several philosophers have let themselves fall into talking at certain times as if they did. (A philosopher in this mood is likely to say, '*X* is *our* notion', with a certain petulance, where *X* may be *reason, truth, justification, evidence, or what have you.*)

This view is, however, self-refuting, at least in our culture. I have discussed this elsewhere (Put-

nam, 1981); the argument turns on the fact that our culture, unlike totalitarian or theocratic cultures, does not have 'norms' which decide *philosophical* questions. (Some philosophers have thought it does; but they had to postulate a 'depth grammar' accessible only to *them*, and not describable by ordinary linguistic or anthropological investigation.) Thus the philosophical statement:

A statement is true (rightly assertible) only if it is assertible according to the norms of modern European and American culture

is itself neither assertible nor refutable in a way that requires assent by everyone who does not deviate from the norms of modern European and American culture. So, if this statement is true, it follows that it is not true (not rightly assertible). Hence it is not true QED. (I believe that *all* theories which identify truth or right assertibility with what people agree with, or with what they would agree with in the long run, or with what educated and intelligent people agree with, or with what educated and intelligent people would agree with in the long run, are contingently self-refuting in this same way.)

Cultural imperialism would not be contingently self-refuting in this way if, as a matter of contingent fact, our culture were a totalitarian culture which erected its own cultural imperialism into a required dogma, a culturally normative belief. But it would still be wrong. For every culture has norms which are vague, norms which are unreasonable, norms which dictate inconsistent beliefs. We have all become aware how many inconsistent beliefs about *women* were culturally normative until recently, and are still strongly operative, not only in subcultures, but in all of us to some extent; and examples of inconsistent but culturally normative beliefs could easily be multiplied. Our task is not to mechanically *apply* cultural norms, as if they were a computer program and we were the computer, but to interpret them, to criticize them, to bring them and the ideals which inform them into reflective equilibrium. Cavell has aptly described this as 'confronting the culture with itself, along the lines in which it meets in me'. And he adds (Cavell, 1979, p. 125), 'This seems to me a task that warrants the name of Philosophy.' In this sense, we are all called to be philosophers, to a greater or lesser extent.

The culturalist, relativist or imperialist, like the historicist, has been caught up in the fascination of

something really fascinating; but caught up in a sophomoreish way. Traditions, cultures, history, deserve to be emphasized, as they are not by those who seek Archimedean points in metaphysics or epistemology. It is true that we speak a public language, that we inherit versions, that talk of truth and falsity only make sense against the background of an 'inherited tradition', as Wittgenstein says. But it is also true that we constantly remake our language, that we make new versions out of old ones, and that we have to use reason to do all this, and, for that matter, even to understand and apply the norms we do not alter or criticize. Consensus definitions of reason do not work, because consensus among grown-ups *presupposes* reason rather than defining it.

Quinian Positivism

The slogan 'epistemology naturalized' is the title of a famous paper by Quine (1969). If I have not discussed that paper up to now, it is because Quine's views are much more subtle and much more elaborate than the disastrously simple views we have just reviewed, and it seemed desirable to get the simpler views out of the way first.

Quine's philosophy is a large continent, with mountain ranges, deserts, and even a few Okefenokee Swamps. I do not know how all of the pieces of it can be reconciled, if they can be; what I shall do is discuss two different strains that are to be discerned in Quine's epistemology. In the present section I discuss the positivistic strain; the next section will discuss 'epistemology naturalized'.

The positivist strain, which occurs early and late, turns on the notion of an *observation sentence*. In his earliest writings, Quine gave this a phenomenological interpretation but, since the 1950s at least, he has preferred a definition in neurological and cultural terms. First, a preliminary notion: The *stimulus meaning* of a sentence is defined to be the set of stimulations (of 'surface neurons') that would 'prompt assent' to the sentence. It is thus supposed to be a *neurological* correlate of the sentence. A sentence may be called 'stimulus-true' for a speaker if the speaker is actually experiencing a pattern of stimulation of his surface neurons that lie in its stimulus meaning; but one should be careful to remember that a stimulus-true sentence is not necessarily true *simpliciter*. If you show me a life-like replica of a duck, the sentence, 'That's a

duck', may be stimulus-true for me, but it isn't true. A sentence is defined to be an *observation sentence* for a community if it is an occasioned sentence (one whose truth value is regarded as varying with time and place, although this is not the Quinian definition) and it has the *same stimulus meaning* for all speakers. Thus 'He is a bachelor' is not an observation sentence, since different stimulations will prompt you to assent to it than will prompt me (we know different people); but 'That's a duck' is (nearly enough) an observation sentence. Observe that the criterion is supposed to be entirely physicalistic. The key idea is that observation sentences are distinguished among occasioned sentences by being keyed to the same stimulations *intersubjectively*.

Mach held that talk of unobservables, including (for him) material objects, is justified only for reasons of 'economy of thought'. The business of science is *predicting regularities in our sensations*; we introduce 'objects' other than sensations only as needed to get theories which neatly predict such regularities.

Quine (1975) comes close to a 'physicalized' version of Mach's view. Discussing the question, whether there is more than one correct 'system of the world', he gives his criteria for such a system: (1) it must predict a certain number of stimulus-true observation sentences;³ (2) it must be finitely axiomatized; (3) it must contain nothing unnecessary to the purpose of predicting stimulus-true observation sentences and conditionals. In the terminology Quine introduces in this paper, the theory formulation must be a 'tight fit'⁴ over the relevant set of stimulus-true observation conditionals. (This is a formalized version of Mach's 'economy of thought'.)

If this were all of Quine's doctrine, there would be no problem. It is reconciling what Quine says here with what Quine says elsewhere that is difficult and confusing. I am *not* claiming that it is impossible however; a lot, if not all, of what Quine says *can* be reconciled. What I claim is that Quine's position is much more complicated than is generally realized.

For example, what is the *status* of Quine's ideal 'systems of the world'? It is tempting to characterize the sentences in one of Quine's ideal 'theory formulations' as *truths* (relative to that language and that choice of a formulation from among the equivalent-but-incompatible-at-face-value formulations of what Quine would regard as the *same* theory) and as *all* the truths (relative to the same

choice of language and formulation), but this would conflict with *bivalence*, the principle that *every* sentence, in the ideal scientific language Quine envisages, is true or false.

To spell this out: Quine's ideal systems of the world are *finitely axiomatizable theories*, and contain standard mathematics. Thus Gödel's celebrated result applies to them: there are sentences in them which are neither provable nor refutable on the basis of the system. If being *true* were just being a theorem in the system, such sentences would be neither true nor false, since neither they nor their negations are theorems. But Quine (1981) holds to bivalence.

If Quine were a metaphysical realist there would again be no problem: the ideal system would contain everything that could be *justified* (from a very idealized point of view, assuming knowledge of all observations that *could* be made, and logical omniscience); but, Quine could say, the undecidable sentences are still determinately true or false — only we can't tell which. But the rejection of metaphysical realism, of the whole picture of a determinate 'copying' relation between words and a noumenal world, is at the heart of Quine's philosophy. And, as we shall see in the next section, 'justification' is a notion Quine is leery of. So what is he up to?⁵

I hazard the following interpretation: bivalence has *two* meanings for Quine: a 'first-order' meaning, a meaning as viewed *within* the system of science (including its Tarskian metalanguage) and a 'second-order' meaning, a meaning as viewed by the philosopher. In effect, I am claiming that Quine too allows himself a 'transcendental' standpoint which is different from the 'naive' standpoint that we get by just taking the system at face value. (I am not claiming that this is *inconsistent* however; some philosophers feel that such a move is *always* an inconsistency, but taking this line would preclude using *any* notion in science which one would explain away as a useful fiction in one's commentary on one's first-order practice. There was an inconsistency in the case of the methodological solipsist, because he claimed his first-order system reconstructed the *only* way he could understand the notion of another mind; if he withdraws that claim, then his position becomes perfectly consistent; it merely loses all philosophical interest.)

From *within* the first-order system, '*p* is true or *p* is false' is simply true; a derivable consequence of the Tarskian truth definition, given standard pro-

positional calculus. From *outside*, from the meta-linguistic point of view Quine occupies, there is no unique 'world', no unique 'intended model'. Only *structure* matters; every model of the ideal system (I assume there is just one ideal theory, and we have fixed a formulation) is an intended model. Statements that are provable are true in *all* intended models; undecidable statements are true or false in each intended model, but not *stably* true or false. Their truth value varies from model to model.

If *this* is Quine's view, however, then there is still a problem. For Quine, what the philosopher says from the 'transcendental' standpoint is subject to the same methodological rules that govern ordinary first-order scientific work. Even mathematics is subject to the same rules. Mathematical truths, too, are to be certified as such by showing they are theorems in a system which we need to predict sensations (or rather, stimulus-true observation conditionals), given the physics which we are constructing as we construct the mathematics. More precisely, the *whole system of knowledge* is justified *as a whole* by its utility in predicting observations. Quine emphasizes that there is no room in this view for a special status for philosophical utterances. There is no 'first philosophy' above or apart from science, as he puts it.

Consider, now, the statement:

A statement is *rightly assertible* (true in all models) just in case it is a theorem of the relevant 'finite formulation', and that formulation is a 'tight fit' over the appropriate set of stimulus-true observation conditionals.

This statement, like most philosophical statements, does not imply *any* observation conditionals, either by itself or in conjunction with physics, chemistry, biology, etc. Whether we say that some statements which are undecidable in the system are really rightly assertible or deny it does not have any effects (that one can foresee) on prediction. Thus, *this* statement *cannot* itself be rightly assertible. In short, *this* reconstruction of Quine's positivism makes it *self-refuting*.

The difficulty, which is faced by all versions of positivism, is that positivist exclusion principles are always self-referentially inconsistent. In short, *positivism produced a conception of rationality so narrow as to exclude the very activity of producing that conception*. (Of course, it also excluded a great many other kinds of rational activity.) The

problem is especially sharp for Quine, because of his explicit rejection of the analytic/synthetic distinction, his rejection of a special status for philosophy, etc.

It may be, also, that I have just got Quine wrong. Quine would perhaps reject the notions of 'right assertibility', 'intended model', and so on. But then I just don't know *what* to make of this strain in Quine's thought.

'Epistemology Naturalized'

Quine's paper 'Epistemology naturalized' takes a very different tack. 'Justification' has failed. (Quine considers the notion only in its strong 'Cartesian' setting, which is one of the things that makes his paper puzzling.) Hume taught us that we *can't* justify our knowledge claims (in a foundational way). Conceptual reduction has also failed (Quine reviews the failure of phenomenalism as represented by Carnap's attempt in the *Logische Aufbau*.) So, Quine urges, let us give up epistemology and 'settle for psychology'.

Taken at face value, Quine's position is sheer epistemological eliminationism: we should just *abandon* the notions of justification, good reason, warranted assertion, etc., and *reconstrue* the notion of 'evidence' (so that the 'evidence' becomes the sensory stimulations that *cause us* to have the scientific beliefs we have). In conversation, however, Quine has repeatedly said that he didn't mean to 'rule out the normative'; and this is consistent with his recent interest in such notions as the notion of a 'tight fit' (an economical finitely axiomatized system for predicting observations).

Moreover, the expression 'naturalized epistemology' is being used today by a number of philosophers who explicitly consider themselves to be doing normative epistemology, or at least methodology. But the paper 'Epistemology naturalized' really does rule all that out. So it's all *extremely* puzzling.

One way to reconcile the conflicting impulses that one sees at work here might be to replace justification theory by reliability theory in the sense of Goldman; instead of saying that a belief is justified if it is arrived at by a reliable method, one might say that the notion of justification should be *replaced* by the notion of a verdict's being the product of a reliable method. This is an *eliminationist* line in that it does not try to reconstruct or analyze the traditional notion; that was an

intuitive notion that we now perceive to have been defective from the start, such a philosopher might say. Instead, he proposes a *better* notion (by his lights).

While some philosophers would, perhaps, move in this direction, Quine would not for a reason already given: Quine rejects metaphysical realism, and the notion of reliability presupposes the notion of *truth*. Truth is, to be sure, an acceptable notion for Quine, if defined à la Tarski, but so defined, it cannot serve as the primitive notion of epistemology or of methodology. For Tarski simply defines 'true' so that '*p* is true' will come out equivalent to '*p*'; so that, to cite the famous example, '*Snow is white*' is true will come out equivalent to '*Snow is white*'. What the procedure does is to define 'true' so that saying that a statement is true is equivalent to *assenting* to the statement; truth, as defined by Tarski, is not a *property* of statements at all, but a syncategorematic notion which enables us to 'ascend semantically', i.e., to talk about sentences instead of about objects.⁶

I will assent to '*p* is true' whenever I assent to *p*; therefore, I will accept a method as reliable whenever it *yields verdicts I would accept*. I believe that, in fact, this is what the 'normative' becomes for Quine: the search for methods that yield verdicts that one oneself would accept.

Why We Can't Eliminate the Normative

I shall have to leave Quine's views with these unsatisfactory remarks. But why not take a full-blown eliminationist line? Why *not* eliminate the normative from our conceptual vocabulary? Could it be a superstition that there is such a thing as reason?

If one abandons the notions of justification, rational acceptability, warranted assertibility, right assertibility, and the like, completely, then 'true' goes as well, except as a mere device for 'semantic ascent', that is, a mere mechanism for switching from one level of language to another. The mere introduction of a Tarskian truth predicate cannot define for a language any notion of *rightness* that was not already defined. To reject the notions of justification and right assertibility while *keeping a metaphysical realist notion of truth* would, on the other hand, not only be *peculiar* (what ground could there be for regarding truth, in the 'correspondence' sense, as *clearer* than right assertibility?), but incoherent; for the notions the

naturalistic metaphysician uses to explain truth and reference, for example the notion of causality (explanation), and the notion of the *appropriate type* of causal chain depend on notions which presuppose the notion of reasonableness.

But if *all* notions of rightness, both epistemic and (metaphysically) realist are eliminated, then what are our statements but noise-makings? What are our thoughts but *mere* subvocalizations? The elimination of the normative is attempted mental suicide.

The notions, 'verdict I accept' and 'method that leads to verdicts I accept' are of little help. If the *only* kind of rightness any statement has that I can understand is 'being arrived at by a method which yields verdicts I accept', then I am committed to a solipsism of the present moment. To solipsism, because this is a methodologically solipsist substitute for assertibility ('verdicts I accept'), and we saw before that the methodological solipsist is only consistent if he is a real solipsist. And to solipsism of the present moment because this is a *tensed* notion (a substitute for warranted assertibility at a *time*, not for assertibility in the best conditions); and if the *only* kind of rightness my present 'subvocalizations' have is *present* assertibility (however defined); if there is no notion of a *limit* verdict, however fuzzy; then there is no sense in which my 'subvocalizations' are *about* anything that goes beyond the present moment. (Even the thought 'there is a future' is 'right' only in the sense of being *assertible at the present moment*, in such a view.)

One could try to overcome this last defect by introducing the notion of 'a verdict I would accept *in the long run*', but this would at once involve one with the use of counterfactuals, and with such

notions as 'similarity of possible worlds'. But it is pointless to make further efforts in this direction. Why should we expend our mental energy in convincing ourselves that we aren't thinkers, that our thoughts aren't really *about* anything, noumenal or phenomenal, that there is *no* sense in which any thought is *right* or *wrong* (including the thought that no thought is right or wrong) beyond being the verdict of the moment, and so on? This is a self-refuting enterprise if there ever was one! Let us recognize that one of our fundamental self-conceptualizations, one of our fundamental 'self-descriptions', in Rorty's phrase, is that we are *thinkers*, and that *as* thinkers we are committed to there being *some* kind of truth, some kind of correctness which is substantial and not merely 'disquotational'. That means that there is no eliminating the normative.

If there is no eliminating the normative, and no possibility of reducing the normative to our favorite science, be it biology, anthropology, neurology, physics, or whatever, then where are we? We might try for a grand theory of the normative in its *own* terms, a formal epistemology, but that project seems decidedly overambitious. In the meantime, there is a great deal of philosophical work to be done, and it will be done with fewer errors if we free ourselves of the reductionist and historicist hang-ups that have marred so much recent philosophy. If reason is both transcendent and immanent, then philosophy, as culture-bound reflection and argument about eternal questions, is both in time and eternity. We don't have an Archimedean point; we always speak the language of a time and place; but the rightness and wrongness of what we say is not *just* for a time and a place.

Notes

1 I chose brown because brown is not a spectral color. But the point also applies to spectral colors: if being a color were purely a matter of reflecting light of a certain wavelength, then the objects we see would change color a number of times a day (and would all be black in total darkness). Color depends on background conditions, edge effects, reflectancy, relations to amount of light, etc. Giving a description of all of these would only define *perceived* color; to define the 'real' color of an object one also needs a notion of 'standard conditions': traditional philosophers would have said that the color of a red object is a power (a disposition) to look red to normal observers under

normal conditions. This, however, requires a counterfactual conditional (whenever the object is *not* in normal conditions). What makes color terms physically undefinable is not that color is subjective but that it is *subjunctive*. The common idea that there is some one molecular structure (or whatever) common to all objects which look red 'under normal conditions' has no foundation: consider the difference between the physical structure of a red star and a red book (and the difference in what we count as 'normal conditions' in the two cases).

2 This argument appears in Firth's Presidential Address to the Eastern Division of the American

Philosophical Association (29 December 1981), titled 'Epistemic merit, intrinsic and instrumental'. Firth does not specifically refer to evolutionary epistemology, but rather to 'epistemic utilitarianism'; however, his argument applies as well to evolutionary epistemology of the kind I describe.

- 3 Quine actually requires that a 'system of the world' predict that certain 'pegged observation sentences' be true. I have oversimplified in the text by writing 'observation sentence' for 'pegged observation sentence'. Also the 'stimulus meaning' of an observation sentence includes a specification of conditions under which the speaker *dissents*, as well as the conditions under which he assents. The details are in Quine (1975).
- 4 A theory is a 'tight fit' if it is interpretable in every axiomatizable theory which implies the observation conditionals (conditionals whose antecedent and con-

sequent are pegged observation sentences) in question in a way that holds the pegged observation sentences fixed. To my knowledge, no proof exists that a 'tight fit' even exists, apart from the trivial case in which the observation conditionals can be axiomatized *without* going outside of the observation vocabulary.

- 5 Quine *rejected* the interpretation I offer below (discussion at Heidelberg in 1981), and opted for saying that our situation is 'asymmetrical': he is a 'realist' with respect to his *own* language but not with respect to other languages.
- 6 Quine himself puts this succinctly. 'Whatever we affirm, after all, we affirm as a statement within our aggregate theory of nature as we now see it; and to call a statement true is just to reaffirm it.' (Quine, 1975, p. 327)

References

- Austin, J. L., 1961. "A Plea for Excuses," in his *Philosophical Papers* (Oxford: Oxford University Press).
- Cavell, Stanley, 1979. *The Claim of Reason* (Oxford: Oxford University Press).
- Firth, Roderick, 1981. "Epistemic Merit, Intrinsic and Instrumental," Presidential Address to Eastern Division of American Philosophical Association.
- Goldman, Alvin I., 1979. "What Is Justified Belief?" this vol., ch. 27.
- Putnam, Hilary, 1981. *Reason, Truth and History* (Cambridge: Cambridge University Press).
- Quine, W. V., 1969. "Epistemology Naturalized," this vol., ch. 23.
- , 1975. "On Empirically Equivalent Systems of the World," *Erkenntnis* 9, pp. 313–28.
- , 1981. "What Price Bivalence?" *Journal of Philosophy* 78, pp. 90–5.
- Rorty, Richard, 1979. *Philosophy and the Mirror Of Nature* (Princeton, NJ: Princeton University Press).
- , 1980. "Pragmatism, Relativism and Irrationalism," *Proceedings and Addresses of the American Philosophical Association* 53.
- Strawson, P. F. 1979. "Universals," in *Midwest Studies in Philosophy*, vol. 4 (Minneapolis: University of Minnesota Press).

The Old Skepticism, the New Foundationalism, and Naturalized Epistemology

Robert Audi

There are many kinds of skepticism. Skeptical positions differ in scope, in modality, in order, in target, and in many other ways. The target may be justification or knowledge or both. It may be claimed that there is no justification or knowledge, or that there *cannot* be any. A skeptic might, however, grant that there *can* be justification or knowledge but deny that we can *know* or be justified in believing that there is any: even if we have first-order justification or knowledge, second-order justification and knowledge may be beyond our grasp. The issues surrounding skepticism are numerous and complex,¹ and I shall address only skepticism about justification.² Much (though not all) of what must be said in answering skepticism about knowledge can be discovered in examining skepticism about justification; but if it should turn out (as I think it might) that skepticism about justification could be answered in a way that does not eliminate skepticism about knowledge, that fact would still be significant. We would presumably establish that it is at least not unreasonable to believe propositions in the vindicated range. In addition to restricting discussion to skepticism about justification, I will leave aside *strong skepticism*, understood as the view that only beliefs of self-evident propositions, or of propositions self-evidently entailed by the former, can be justified (and I will ignore radical forms of skepticism, which are even stronger). My interest will be in *moderate skepticism*, in

particular a broadly Humean skepticism that allows not only the possibility of justified beliefs of logical truths and of certain propositions self-evidently entailed by them but also of justified beliefs about one's own mental states. It denies, however, that we are justified in believing anything about the external world.³

I An Appraisal of a Cartesian Case for Skepticism about External World Beliefs

I have just referred, as many others have, to "answering" skepticism. But this notion can easily obscure a distinction fundamental in appraising skepticism: the distinction between rebutting skepticism and refuting it. Let us speak of rebutting skepticism when we refer to simply showing that one or more skeptical arguments is not sound or that a skeptical conclusion has not been established, e.g. to showing the invalidity of an argument for the view that there is no justified belief (or, more broadly, to showing the absence of any good reason to believe that skeptical thesis). By contrast, refuting skepticism is showing that a skeptical thesis is *false*, where this implies (assuming the skeptical position is consistent) showing a positive result such as that there is (or at least can be) justification for beliefs about the external world. Rebuttal of a skeptical thesis would entitle us to withhold it, and a rebuttal based on successful criticism of a sufficiently wide range of skeptical arguments might warrant regarding skepticism as, say, probably false; refutation of a skeptical

Originally published in R. Audi, *The Structure of Justification* (Cambridge: Cambridge University Press, 1993), pp. 353–61, 364–72, 375–7.

thesis would entitle us to deny it. Rebuttal is less difficult to achieve than the latter; and suspending judgment on a skeptical claim requires less justification than rejecting it. Both must be considered in appraising “answers” to skepticism.

If any one skeptical narrative epitomizes the challenge of skepticism, it is probably Descartes’s evil demon scenario. Some may feel its power more keenly in its new, naturalized version, the case of the brain in a vat. But the incubus on epistemology is essentially the same. We are still haunted by Descartes’s nightmare. Here is one way to see why. Suppose, to take what seems a paradigm of perceptual justification, that

- (1) I am justified in believing p – that there is a bespeckled black-and-white surface before me.
- (2) The proposition that p , self-evidently entails that there is no demon causing me to have *just* the kind of sensory experience I am now having *without* there being such a surface before me. But
- (3) Necessarily, if I am justified in believing p , I am justified in believing any proposition that self-evidently follows from p ; and
- (4) I am not justified in believing that there is no such demon. Hence,
- (5) I am not justified in believing that p .

A key premise for (4) is that I cannot be justified in believing that there is no such demon when, if there should be one, my *evidence base* would be just what it is now. Nothing in my experience (or otherwise accessible to me) would discriminate between the veridical case in which I see the surface and the demonic case in which I merely hallucinate one.⁴

Skeptical arguments quite similar to this have been widely discussed.⁵ Many philosophers have argued that we are of course justified in believing that there is no such demon. Fewer have argued that (3), the crucial epistemic principle – *the transmission principle*, for short – is false. I think both approaches are promising but will first pursue the latter. The transmission principle is open to prima facie counterexamples. In proposing such an example, I use a strategy of rebuttal; attacking (4) requires attempting a refutation of skepticism: showing that we actually have justification for a belief we hold. I leave that more complicated strategy for Section III.

Consider this possibility. I prove p , a theorem in the propositional calculus, check my results, and

thereby come to be justified in believing that p . Because I sometimes make mistakes and need p for my work, I plan to ask a colleague, who is a logician and whom I justifiably believe to be better at such deductions, to check my work. Suppose that, feeling both my need for p and my invigorating success in my deduction, I now think of p with a sense of surety about it and infer from p that if my colleague says it is false, he is wrong. From the proposition that p , it certainly follows that if he says p is false, he is wrong. If it *is* a theorem (as we are assuming), then if he denies it, he is wrong (i.e., states a falsehood). This indeed seems self-evident, in the sense (roughly speaking) that (a) understanding it is sufficient for being justified in believing it and (b) believing it on the basis of understanding it is sufficient for knowing it.⁶ But even though I am justified in believing that p , am I *automatically* justified in believing the *further* proposition that if he says that it is false, he is wrong? Suppose my checking my proof to the extent I did is only enough to give me the *minimum* basis for justification in believing p . Surely I would then not have sufficient grounds for the further proposition that if he says p is false, he is wrong.⁷ If I had done the proof in two quite different ways and triple-checked each procedure, things might be different.

This minimality assumption is essential: I do not claim that there is *no* degree or kind of justification for which the relevant sort of transmission principle will hold, only that it fails for justification in general, particularly as taken to be such that (a) having it “entitles” one to believe (or warrants one in believing) p and (b) believing a true proposition with that degree of justification implies – apart from such untoward cases as the post-Gettier literature has unearthed – knowing it. That degree of justification, I maintain, is such that I can possess it on the basis of a careful deduction of p and yet not be justified in believing all of p ’s self-evidently entailed consequences. This degree can be quite high; my case requires only that, whatever the appropriate standard, it be just minimally met. If I had begun with indefeasible justification, then perhaps I would be justified in believing any self-evidently entailed consequence – at least if I could see that it was entailed.⁸

A crucial point here – and one that may go some distance toward explaining why we should not expect unrestricted closure principles for justification – is that the preservation of truth need not be precisely paralleled by the preservation of justification. Truth is an ontic notion, justification an

epistemic notion; and we may not simply assume that logical principles, which formulate conditions for carrying truth into truth (i.e., for validity), are mirrored by epistemic principles. It is possible that Descartes foresaw that epistemic transmission is not precisely parallel to validity. Wanting to build a “firm and permanent structure in the sciences,” he insisted on an epistemic strength in the foundations so great that it could not be attenuated by deduction and would reach as far as valid inference might take us.

There is a great deal to be said about the epistemic principle in question: viz., that justification is automatically transmitted across self-evident entailments. There is no hope of doing justice to the topic here.⁹ It should help, however, to set forth weaker transmission principles that may hold and may partly explain the appeal of the strong one that so powerfully aids skepticism. Nothing I have said is inconsistent with the principle that

- I. If there is reason (in the form of one or more reasons) to believe p , then there is reason to believe any proposition self-evidently entailed by p (hence to believe false any proposition self-evidently entailing not- p) – call this *the closure of reasons principle*.

A relativized parallel would be this:

- II. If there is reason for S to believe p , and p self-evidently entails q , then there is reason for S to believe q .

It is of course S 's *having* reason to believe p that is closest to S 's being justified in believing it. A parallel principle for the former notion is this:

- III. If S has reason to believe p , p entails q , and S can understand both q and the entailment of q by p , then S has reason to believe q – call this *the principle of closure for having reasons*.¹⁰

By way of clarification, I want to make four points. First, it is possible to have a reason for believing p while at the same time having a reason, even a better reason, for believing not- p . Second, the principle of closure for having reasons implies that such reasons are *indefinitely extendable*: they reach as far as entailments one can understand. If, for instance, one has a reason for q in virtue of one's having a reason for p and understanding its entailment of q , and the same holds for q in relation to r , then one has a reason for r ; and so on. We can

thus distinguish between having an *immediate* reason for believing something, which occurs where one has a reason for it not transmitted from a prior reason, and having a *mediated* reason for believing, which is a reason one has for it that *does* arise by transmission from a prior reason. However – and this is the third point – as the first two points imply, the strength of the reason that is transmitted across the entailment may diminish progressively and approach zero. Finally, although self-evident entailment will often be the connective tissue in transmission of reasonhood, indefinite extendability is neither equivalent to, nor necessarily due to, transitivity of self-evident entailment. It is in fact not clear whether that relation is transitive.¹¹

Distinguishing the principle of closure for having reasons from the transmission principle is of the first importance for the transmission issue. The principles are quite different; but they are easily conflated, partly because having a reason for p implies having *some* degree of justification for it. I have not denied any of these closure principles for reasons; indeed, I think they are plausible. I suspect it is the plausibility of weaker principles like these that makes the stronger transmission principle look so plausible when it is not distinguished from the others.

To see the difference between the transmission principle and the principle of closure for having reasons, recall that one can have *some* degree of justification for believing p yet not be justified, on balance, in believing it. Similarly – and, on some views, equivalently – one can have *a reason* (or some reason) to believe p , without being justified, overall, in believing it: without, we might say, having *adequate* reason to believe it. To reduce the vagueness of this formulation, we can regard the crucial level of justification relevant in discussions of skepticism – which we might simply call *adequate justification* – as the degree such that, when there is no Gettier-type problem, then p , if true, is known.¹² Alternatively, but I think not equivalently,¹³ we can speak of that degree of justification sufficient to entitle an epistemically rational agent to believe p .¹⁴ People may differ concerning what degree (or kind) this is. My point is that it is scarcely controversial that this status is not achieved by merely having *some* degree of justification for believing p , or merely *some* reason to believe it, any more than a task with many parts is completed by finishing just one of them.

The overall view I presuppose here is that the degree of one's justification for p , like one's justification simpliciter for believing p , is determined by one's overall relevant epistemic situation at the time. Many factors are pertinent, and they include not only the number and strength of defeaters, both counterevidence and reasons to doubt the efficacy of one's justifiers. To suppose that my epistemic situation relative to q must be the same, or at least as good, as my epistemic situation relative to p , simply because p self-evidently entails q and I am justified in believing p , is a manifestation of a simple *linear* concept of justification. The picture has justification traversing a line from a belief that has that justification to one that as yet does not. This omits – for one thing – the relevance of *adding* q to my system of beliefs (or other cognitions). That addition provides a new epistemic context, and the offspring may quite surprise the parents.

How do these reflections bear on answering skepticism? Perhaps we have rebutted the evil demon argument (or one such argument), which apparently depends on some similarly defective transmission principle; but if the weaker transmission principle, for having reasons, should hold, do we not face another Cartesian nightmare? For is there any better ground for holding that I have *some* reason to believe no demon is deceiving me than for holding that I am (overall) justified in believing this? I think there is; at least, there is *if* I have *some* reason to believe that there is a bespeckled surface before me, which seems quite evident, and if I may also presuppose something like the principle of closure for having reasons. But certainly we may say this much: we are at least better off if the task is to produce only some reason to believe p rather than overall justification for believing it. The incubus is not as heavy, however much we may still dislike the burden. But how might we escape that burden, and can foundationalist theories be of any special help?

III An Approach to the Refutation of Skepticism

I have stressed that, at least if a foundationalist approach to justification is sound, an inability to show that one is justified in believing something by no means entails that one is not justified therein. Inability to produce a justification of a belief does not entail that I lack one. Indeed, the

objects in the two cases are different. The justificatory process I would engage in would produce second-order justification, justification for the proposition that I am justified in believing p (or at least that p is justified); the justification skeptically doubted in the first place is simply *for believing the first-order proposition that p* (or, perhaps, for p). This is easy to forget in the dialectic with skepticism. For the skeptic challenges our common-sense belief that we have a great deal of justification; and we want not just to rebut the arguments to this effect but also to refute their conclusions: not just to show that the skeptic's case is unsound but also to show that various kinds of beliefs we have are justified. But even epistemologists who realize that one can be justified in believing p without one's being able to show that one is are not satisfied with this point; there would be a historical imperative to refute skepticism even if epistemology could be plausibly held to be purely theoretical and not to depend on existential propositions to the effect that anyone actually has justification. Can this imperative be satisfied for anything so substantive as fallible beliefs of contingent propositions?¹⁵

Much depends, of course, on what counts as showing something (a topic that deserves a full-scale study in itself). Skeptics will naturally hold that only *entailing* grounds for p will show it; and although I deny this – as begging the question against the cogency of inductive reasoning – I shall try to use only deductively valid arguments in assessing the case for showing that there is justification. We must distinguish at least two ways of showing that we have justified belief: (1) *dialectical showing*, the second-order process of showing, by subsumptive appeal to an epistemic principle, that we have a justified belief, as where one subsumes a kind of belief under a generalization to the effect that all beliefs of that kind are justified; and (2) *simple showing*, the first-order process of simply *giving* a justification for a belief, in the sense of laying out that justification by adducing an adequate justifying ground for p in a way that indicates that p is justified, but without making any second-order claim. To illustrate simple showing, suppose I am asked to show that I am justified in believing that there is a bespeckled white surface before me. I might say that I see it or, pushed a bit by reminders of possible hallucinations, that I have a clear visual impression of such a surface. If this impression *does* in fact justify me, then I have given a justification: I have laid it

out in a way that in some sense shows that I am justified, as, by exhibiting my pen, I might show that I have one.

This demonstration might be claimed to be merely a case of *exhibitive showing*: showing *what* justifies, not of (propositionally) showing in any sense *that* one is justified. I prefer to say that it is showing *that* one is justified by showing *what* justifies one; but if one accomplished only the latter, it still provides the *basis* for showing, in some way, that one is justified, e.g. by reflecting on how the exhibited ground is related to *p*, and thereby discovering a crucial epistemic premise usable in dialectically showing that one is justified. I grant, however, that in the example imagined I certainly have not used a second-order concept or appealed to an epistemic principle. Even the moderate skeptic wants nothing less than this; and – rightly or wrongly – skeptics will not concede that one has in any sense shown a justification unless one has done so dialectically.

If we set out to do that, we might ask what sorts of epistemic principles we may use to back up the claim that a belief we hold is justified and how we might be justified in believing *them*. The principle suggested by my example is something like this, understood as applying to all the senses:

IV. If S has a clear sensory impression of *x*'s being *F* (or that *x* is *F*) and on the basis of that impression believes that *x* is *F*, then this belief is (prima facie) justified – call this *the perceptual principle*.¹⁶

To see how this principle can be used against skepticism we need a further distinction: between *personal showing*, the kind done by people, and *impersonal, argumental showing*, the kind done by what we call cogent arguments; impersonal showing, in turn, can be either simple or dialectical. Personal showing is a process; impersonal showing is a kind of epistemic status of an argument. An argument impersonally shows that *p* when its premise set provides an adequate ground for *p*, where adequacy may be either relativized to evidence accessible in the context, for instance to mathematical or scientific or sensory information, or taken more broadly, as implying only that the appropriate grounds *exist*. (Adequacy needs further analysis, to be sure – e.g., there are different cases depending on whether or not the ground can be *seen* by some relevant person or procedure to be one.) A person, S, shows that *p* when S provides an adequate ground for *p*; and if S is justified in

accepting that ground and in using it to try to show that *p*, we may say that S personally (in a sense, subjectively) shows *p*, since the result is to produce justification for S's believing it.¹⁷ It is personal showing that is of greatest interest here: we want to know whether we can show certain important skeptical claims to be false, not just whether there are (or can be) arguments that show this.

If I am to show that my belief that there is a bespeckled white surface here is justified, how can I be warranted in believing an epistemic principle like the one in question? Let us first consider the possibility that one can know a priori that such sensory impressions yield prima facie justification for perceptual beliefs of the kind at issue. One argument for this would be that such principles are partly constitutive of the concept of justification; hence they are justifiably believable on the basis of conceptual reflection, e.g., reflection about what one has to count as a justifier in order to qualify as having the concept of justification. Granted, such reflection would presumably have to range over cases of belief in which justification can be concretely (and intuitively) seen; for instance, we might try to imagine how someone could (i) understand the concept of justification yet (ii) sincerely claim, of my belief that there is a bespeckled white surface before me, that even though it is based on my clear visual impression of such a surface, it is not prima facie justified. It is at best difficult to imagine how (i) and (ii) can jointly hold. This confirms the status of the perceptual principle as partly constitutive of the concept of justification. These concrete cases we use to illustrate such an epistemic principle might be hypothetical products of a priori reflection itself; they need not be an inductive evidence base for the principle, as opposed to concrete applications of it. Roughly, it is largely because we grasp the principle that we can construct the cases in question; we do not acquire the principle – as opposed to a formulation of it – by first constructing the cases. If the cases do constitute a kind of inductive base, they need not function in the way observations do in enumerative arguments for contingent generalizations.¹⁸

To be sure, it could be argued that such principles are not only a priori, but analytic, roughly in the sense that we can arrive at them directly by analysis of an appropriate concept. Suppose that justified belief can be analyzed as, simply, belief that accords with at least one of a certain set of epistemic principles, where the set includes the

perceptual principle. Then the principle will be true “by definition.” This interesting possibility deserves more discussion than I can give it here. I have no need to reject it, but I will make just one point to suggest why I resist it. Even if we could find a set of epistemic principles comprehensive enough – and sufficiently free from even preanalytic disagreements – to constitute a plausible analysis, there would remain the question of the status of our warrant for accepting *it*. I do not think, e.g., that its truth could be known on the basis of empirical linguistic knowledge about “justification.”¹⁹ There might indeed be as good reason to think such an analysis synthetically true as to think this of its constituent epistemic principles.

It is a further question whether there can be *non-inferential* justification for such epistemic principles. It is not self-evident that inductive grounding is the only kind of inferential justifiedness they might have; hence, ruling that out leaves other kinds of inferential justification open. But notice that, for the foundationalism introduced here, the non-inferential justification of epistemic principles is consistent with the possibility that they are (1) inferentially justifiable, (2) justifiedly believed only after considerable reflection, and (3) defeasibly justified even then. They may be inferentially justifiable because foundational status does not imply *ultimacy* in the epistemic order, in the Aristotelian sense that there is nothing “prior” to *p*.²⁰ They may be justifiably believable only after reflection because the required kind of self-evidence is a matter of the kind of non-inferential knowledge obtainable by understanding, not a matter of the ease or speed with which the understanding comes; and given the fallibilism of the approach in question, justified beliefs of these principles may be defeasible. Thus, one can have non-inferential justification for the relevant epistemic principles even if they are not self-evident in the way axioms are. They may in fact seem self-evident only after long reflection on appropriate examples; that helps to explain how they can be defeasibly justified, but it does not imply that their justification is inferential.

Now I think that I am justified in believing our perceptual principle, and I will assume that my justification is non-inferential (though that assumption is not essential to the argument so long as I could have an inferential justification of sufficient strength, as seems possible). Suppose that (a) I justifiedly believe both the principle and

that I now have a clear impression of a bespeckled white surface before me and, (b) on that basis, believe there is such a surface before me. From these two propositions it self-evidently follows that I have a justified belief that there is a bespeckled white surface before me. May I not, then, justifiedly believe that I have a (prima facie) justified belief about the external world? And in so arguing, have I not shown, dialectically, that the relevant kind of skepticism is false? On one plausible interpretation of what it is to show something, if one is justified in believing one’s (true) premises, and validly (and non-circularly) argues from them to a conclusion, one has shown it.

I have put my conclusion rhetorically for good reason: I am trying to *exhibit* my dialectical showing and not to *claim* it. Suppose I claim categorically that I have shown that I have a justified belief; the proposition at issue now, if the skeptic challenges me, is the *third-order* one just asserted: that I have shown (hence justified) the second-order proposition that I have a justified belief that *p* (the first-order proposition). I have presumably not shown this third-order proposition; but surely I do not need to in order simply to *have* shown the second-order proposition that I have a justified belief. This point is crucial to understanding the dialectic with skepticism: because we imagine a dialogue, we are always anticipating the skeptic’s questioning whether our responses to a given challenge succeed; and because skeptics have an iterative habit of mind, they tend to query the credentials of everything one asserts. But a reply that shows justification at one level – and in one context, we might say – need not show it at the next, which would indeed require epistemic argument at that higher level (and so in a different context). Giving a justification may answer a request for one but need not show that the request *has* been answered; showing that one has justification establishes that one does, but it need not show that one *has* established this, even if one can do so; and so forth. We must not allow the skeptic’s hunger for ever higher levels of justification to undermine an argument for justification at any level. On my approach we can in principle always move to a higher level; e.g., using the (a priori) principle that if one subsumes a belief (justifiedly) under an appropriate epistemic principle which says that beliefs of this sort are justified, one can, for any justified belief, find an epistemic principle by appeal to which one shows that the belief is justified. Still, our achievements at any one level

do not depend on our having already done so, or even on our being able to do so.

A skeptic who accepted all I have said so far might still claim that unless my premises are self-evident, I must be able to say something on their behalf. Are they? It is arguable that the epistemic principle is self-evident in some sense; and although the instantial premise, that I have a clear impression of a bespeckled white surface, is not, moderate skeptics have tended to allow that we are justified in such occurrent mental state beliefs. Perhaps, however, one may not assume, without inductive evidence, that an impression is *clear*. I think there is a sense of “clear,” strong enough to sustain our epistemic principle, for which one may believe this without inductive evidence; but this point is admittedly not self-evident. Perhaps, however, the wider epistemic principle we get by deleting the requirement of a clear impression is also true – but simply generates less justification. Perhaps when a sensory impression of *x*'s being *F* is not clear, it simply generates less justification for believing *x* is *F* than it would yield if it were clear. Granted, a sufficiently *unclear* impression of *x*'s being *F* may fail to produce justified belief to that effect; but this may be because *unclear* is a defeating feature and not because the presence of an impression of *x*'s being *F* is in itself not sufficient for some degree of justification. In any case, I think we may conclude that principles like the perceptual one may provide some basis for our justifiably believing that we have justified beliefs about the world, and that such principles are, for at least one plausible notion of showing, a basis for rejecting skepticism concerning these beliefs.

At least one serious problem remains, however, on the plausible assumption that a belief not *based*, in a partly causal sense, *on* its justifying ground is not justified *by* that ground.²¹ Am I justified in believing that my belief that there is a bespeckled white surface before me is *based* on my impression to that effect? This is, after all, a partly causal proposition, and, arguably, I need inductive grounds for the relevant justification. Again, we confront a Cartesian nightmare: the demon might sever the causal connection between my belief and its ground, leaving my experience just as it is.²² What justification do I have, then, for believing that the connection holds? Note that this demonic possibility does not threaten the view that we can rebut the skeptic. Skeptical arguments can still be shown to be unsound. But the possibility does

threaten the view that we can dialectically show that skepticism is false in ways we might like. Granting that if I have an adequate experiential ground for *p*, I *can* justifiably believe *p*; still, if the demon can sever the crucial connection, the problem is to show that I *do* justifiably believe it.

Here are two points in reply. First, while I am fallible about whether a perceptual belief of mine is based on a matching impression I have, there is no reason to think that I am generally unjustified in my self-ascriptions of this basis relation; and we surely need not accept any skeptical argument from mere fallibility about *p* to lack of justification for believing it.²³ Indeed, most skeptics are willing to allow that we are justified in believing that we *have* beliefs; raising the question of whether we have any justification for our beliefs in fact *presupposes* that we do. Granted, radical skeptics can conditionalize, formulating all their arguments on the supposition that we have beliefs. This strategy may, in the end, deprive them of some of the materials that make their view seem threatening (e.g. by calling attention to the disparity between what they presuppose themselves – that they are making a case – and what they allow others to presuppose), but let that pass. My second point is unaffected: even if my belief is not based on the ground I have for it, in having that ground I still *have* a justification for *p*. If I can show that I have it, and the skeptic cannot provide good reason to doubt that I believe *on* that ground, my dialectical position is at least tolerable. Although it would be better to be able to show not only that I have justification *for* my beliefs, but also justified beliefs – that I am not only justified *in* believing, but also *justifiably believe* – it is still important to be able to show that I have justification *for* them. One might say that at least I don't believe more than I am entitled to, even though I can't show that I don't believe on a basis other than my entitlement.

If the internalist view taken here is not reliabilist, might it still be naturalistic? Let us continue to distinguish between a substantively naturalistic epistemology and a conceptually naturalistic one. The former takes epistemological propositions, such as epistemic principles, to be empirical; the latter simply uses no irreducibly normative concepts. The former is naturalistic in substance – taking epistemology (as Quine does) to be about the natural world in the same way empirical science is; the latter is naturalistic only in concept, allowing that not all the truths of epistemology are about the natural world. The inspiration for the first view is

presumably the idea that the truths of nature are the only truths there are; the second view simply says that, ultimately, all truths are expressible without relying on normative concepts. Although my view is not substantively naturalistic, I have said nothing here that rules out its being conceptually naturalistic. I do not believe, however, that justification is analyzable in terms of reliability. But I have left open that, in some other way, it might turn out to be a natural property. I do think it supervenes on natural properties and is, in that sense, *rooted* in the natural world, but it is by no means clear to me either that justifiedness is a natural property or that the notion of justification is reducible to some set of naturalistic concepts.

My view may seem anti-empirical, but it is not and is indeed devised with an eye to understanding scientific justification. The view is, however, anti-empiricist, if empiricism is understood as denying that anything substantive can be known or justifiably believed a priori. There is no question that experience is crucial not only for concept

formation but also for determining how fruitful a philosophical theory is. Even if an epistemic principle is a priori and necessary and hence not refutable by experience, the course of experience may be such that the principle ceases to hold interest for us, and we evolve toward a different framework using different concepts. Imagine, e.g., that we ceased to have sensory impressions; there would then be no point in working with perceptual epistemic principles. The epistemic principles we rely on in our daily critical practice seem to me so general, however, and so rich in ways to accommodate ostensible disconfirmation, that I see no reason to expect any such outcome. On the other hand, if we posit only empirical principles, whose justification is at the mercy of contingent events, there is surely less hope of adequately dealing with skepticism. Principles whose justification depends on the specific course of experience seem unlikely to enable us to show that we can have justification no matter how that experience goes and even if, unbeknownst to us, it is plagued by a Cartesian nightmare.

Notes

- 1 The literature is also immense. For a recent survey article see Ernest Sosa's wide-ranging study "Beyond Skepticism, to the Best of Our Knowledge," *Mind* IIIc (1988).
- 2 A number of philosophers think that justification is the more important concept. For an indication of why this might be so, and indeed of the stronger position that the notion of knowledge is not of great philosophical importance, see Mark Kaplan, "It's Not What You Know That Counts," *Journal of Philosophy* LXXXII (1985).
- 3 A Humean skepticism, being empiricist, would deny that there are synthetic entailments, and even a non-Humean skepticism might deny that they are ever self-evident; but it is not obvious that a moderate skepticism, as such, need deny either point. By contrast, any moderate skeptic is likely to allow that simple analytic propositions are justifiably believable, at least on the assumption that we have a priori knowledge of the conceptual equivalences through which analytic propositions can be formally reduced to logical truths through the use of definitions.
- 4 This phenomenon seems equivalent to what Michael Williams has called the neutrality of experience. For some elaboration of his view developed in criticizing Davidson, see his "Skepticism and Charity," *Ratio* (New Series) 1 (1988).
- 5 Sosa, "Beyond Skepticism," gives many references. Particularly relevant are Fred Dretske's "Epistemic Operators," *Journal of Philosophy* 67 (1970), and Peter Klein's *Certainty* (Minneapolis: University of Minnesota Press, 1981). Crispin Wright's "Skepticism and Dreaming: Imploding the Demon," *Mind* C (1991) exhibits still another approach. For extensive criticism of that paper see Anthony Brueckner, "Problems with the Wright Route to Skepticism," *Mind* CI (1992).
- 6 This formulation is taken from my "Moral Epistemology and the Supervenience of Ethical Concepts," *Southern Journal of Philosophy* XXIX Supplement (1991), where it is briefly elaborated. Notice that the characterization does not entail something usually associated with self-evident propositions: being very readily understood by normal adults. Some such requirement could be added here without affecting my arguments. Moreover, I am referring to "fully" understanding, as opposed to partially understanding, *p*; but full understanding does not require *perfect* understanding.
- 7 This example is styled after one I gave in *Belief, Justification, and Knowledge* (Belmont, CA: Wadsworth Publishing Co., 1988). The example has been widely discussed and is critically assessed in Catherine M. Canary and Douglas Odegard, "Deductive Justification," *Dialogue* XXVIII (1989). For further

- discussion, and my reply to them, see my "Justification, Deductive Closure, and Reasons to Believe," *Dialogue* XXX (1991).
- 8 I disallow maximal, in the sense of indefeasible, justification, because the failure of transmission illustrated poses some threat to the initial justification even on the assumption that transmission does not always occur. For as suggested below, since reason-hood apparently does transmit over the relevant entailments, if S can (as in my example) deduce consequences he has reason to believe false, it would seem possible that, given such reasons to believe p false, S might go on to deduce enough disconfirming consequences to yield collective reason sufficient to defeat the original justification for p – in which case it is not indefeasible. An even stronger qualification would be to require that the initial justification must not be *absolute*, where absolute justification for p implies the impossibility *both* of defeat and of there being a reason for believing not- p ; but I see no reason to go this far.
 - 9 In a paper in progress I discuss a wide range of the relevant issues; here I can stress only the holistic character of justification as a general explanation of the failure – indicated by the theorem example – of what might be called simple linear transmission.
 - 10 The reference to understanding here avoids the implication that one can have a reason to believe a proposition one cannot understand or have a reason one does possess transmit over an entailment one cannot understand. One might reply that in either case one has a reason but simply cannot see it as one. Indeed, one could argue that S might fail to understand even a self-evident entailment and hence propose that principle II should, like III, require understanding of the entailment. I leave this open. These points may simply show that there are different notions, or at least different degrees of implicitness, of "having" a reason. There are many problems in formulating transmission principles, and I believe that for different plausible principles we get different notions of what it is to have a reason.
 - 11 It is arguable that if p self-evidently entails q and q self-evidently entails r , then a bit of reflection is all that is needed to see that the first entails the third, and hence the relation is transitive. But first, it is not clear that this follows. A person considering just p and r might need *imagination* to see the intermediate step. Second, even if mere reflection on the relevant propositions would yield a grasp of the intermediary, *much* reflection might be needed. At most, then, we could conclude that what might be called *mediate self-evidence* is transitive. For some discussion of that notion of self-evidence and its importance in metaethics, see my "Moral Epistemology and the Supervenience of Ethical Concepts."
 - 12 At least one further qualification may be needed. Consider the lottery case in which S justifiably believes that S's ticket, being one of a million in a fair lottery, will lose. Arguably, there is neither a Gettier problem here nor a degree of justification that is achievable by simply increasing the number of tickets and sufficient to render S's belief knowledge if it is true.
 - 13 Not equivalently because, in lottery cases, it is not the degree, but the kind, of one's justification that prevents one's belief that one will lose from constituting knowledge – at least this is plausible. I have defended it in ch. 7 of *Belief, Justification, and Knowledge*.
 - 14 I have in mind belief simpliciter, as opposed to belief at a given level of confidence. Confidence levels are relevant and can be taken into account, but doing so here would take us too far afield.
 - 15 I refer just to fallible beliefs of contingent propositions, as opposed to infallible ones such as that I exist, because the latter are not really challenging cases for the skeptic.
 - 16 Here "has the impression that x is F " does not entail believing that x is F but only being disposed to believe that it is (it is also not entailed by believing that x is F , since that is quite possible without having any sensory impression); and in "an impression of x 's being F " the position of " x " is not referential: there need be no such entity, even at the level of sense-data. There are problems about how the content of the impression must be related to the belief, but these need not be settled for our purposes here. Note also that the principle can account for the justification of a belief grounded in a mere hallucination, though such a belief is not perceptual in the strict sense presupposing the occurrence of a normal perception.
 - 17 Not necessarily a justified belief, however, since S could still believe that p for some other, inadequate reason, as opposed to the good reason adduced. Note also that we would allow that S can show that p without being justified on either count, though I think that on reflection we *might* then prefer to say only that S's argument showed that p .
 - 18 Granted, one could take a more particularist approach on which the primary a priori intuitions concern particular instances of the concept of justification. One can also combine particularist and generalist strands and appeal to a strategy of integrating intuitions of general principles with those of relevant particular cases, with an eye to reflective equilibrium. In that case one would presumably argue directly against the skeptical attack on the perceptual beliefs in question. Note, however, that one might still argue a priori for epistemic principles, *provided* one could find a suitable premise, such as that the cases one is considering are representative, which would enable one to deduce

- such principles from what one intuitively about these cases.
- 19 For some reasons to hold this, see *Belief, Justification, and Knowledge*, ch. 4.
 - 20 See *Posterior Analytics* 72b. Descartes may well have been influenced by this strongly axiomatic picture.
 - 21 Though controversial, this assumption is widely accepted.
 - 22 This, I take it, is the sort of worry that troubles Crispin Wright, "Skepticism and Dreaming" and it may be a reason why Chisholm and other internalists never impose any causal requirement on justified belief. He does not approach it in the same way, however.
 - 23 This may of course be questioned; for some points by way of justification for it see ch. 9 of *Belief, Justification, and Knowledge*.



PART VII

Epistemic Externalism

Introduction

It is perhaps useful to think of the varieties of epistemic externalism as denying one or more *internalist* theses. What unifies internalist theses is the thought that knowledge (justification in believing) that *p* is *not* determined solely by facts external to the awareness of the believer. The believer must have some appreciation of the facts in virtue of which she knows, whether these be facts about the reliability of her source of belief, facts about what she would have believed under various conditions, or perhaps only facts about what her reasons are for believing as she did. Talk of determination is somewhat vague, of course. One rightly asks, is determination here to be understood as conceptual reduction or as non-reductive necessitation? But the guiding thought of internalists is that knowledge isn't purely a one-level affair: it requires knowledge (justified belief) about that knowledge. To use an example of Keith Lehrer's: Truetemp's merely having a reliable mechanism in his brain for producing accurate highly discriminatory beliefs about the temperature doesn't by itself suffice to make any of his correct temperature judgments reasonable, let alone pieces of knowledge.

In his selection, Alvin Goldman formulates an externalist theory of justification, according to which a belief is justified if and only if it is produced by a reliable belief-forming process. This is eventually modified to take account of the difference between processes that take beliefs as inputs (belief-dependent processes) and those that do not (belief-independent processes). Thus, justification in the end is specified recursively. Goldman sees the principal intuitive benefit of reliabilism to be that it accounts for the relevance to justification of

the quality of the way the belief was actually produced. Wishful beliefs and hasty judgments lack justification because they are the products of untrustworthy ways of believing.

Goldman is also aware of several possible problems for his account, to each of which he proposes solutions, some more tentative than others. Here are several of the problems his account faces. Suppose wishful thinking were reliable in some world unlike ours. Would its products in such a world count as justified? This problem has a much-discussed counterpart: Is the victim of a Cartesian evil demon totally lacking in justified belief merely owing to bad luck in having unreliable faculties of perception, testimony, and the rest? Again, it seems that someone who has excellent undermining evidence for a belief that, as it turns out, is reliably produced cannot be justified in believing as he does, and yet reliabilism has it that he is justified.

Perhaps the most fundamental problem facing Goldman's account is the generality problem. The most recent and sustained discussion of this problem, which defends reliabilism, is by William Alston, in the selection included here.

Attempted reliabilist solutions of the generality problem, especially Alston's, are discussed and criticized in detail by Conee and Feldman in their selection. If reliabilism is to give correct treatments of ordinary cases, processes must be individuated neither too broadly nor too narrowly. If the individuation is too broad, and the process is something of the order of *inferring that p from one's mentioning that p*, then the process is likely to be unreliable, and so its products wrongly counted as unjustified in every case. If it is too narrow, the

process will close in on the particular few or one product(s) produced, so that if these should turn out to be lucky guesses, then they wrongly count as justified.

Keith Lehrer and Richard Fumerton offer internalist objections to externalism. Both cite its naturalistic character as the source of its implausibility. Lehrer argues that natural relations are not sufficient for a subject's having the proper background information about his beliefs and their reasonableness, a claim well-supported by the case of Mr Truetemp. To have knowledge and justification, one's belief must cohere with background information about one's trustworthiness.

Fumerton argues that if some form of externalism were correct, normative epistemology would not lie within the province of philosophy. Scientists, rather than philosophers (qua philosophers), have the qualifications necessary for identifying and describing the causal/nomological features of beliefs with which externalists, according to Fumerton, identify epistemological features such as justification. But this is not Fumerton's principal criticism in the end. He objects most of all to externalism's consequence that there is nothing epistemically problematic in using a method of belief to show that it is reliable. Given externalism, there simply wouldn't be a problem of epistemic circularity. Fumerton claims there is no philosophically interesting conception of justification that meets this description, a fact revealed by the impossibility of even getting skeptical arguments off the ground in an externalist framework. Higher-order questions of justification, under externalism, cannot be treated any differently than lower-order questions.

John McDowell attempts to give a view of knowledge that avoids the pitfalls of coherentism and externalism. What is right in coherentism is the Sellarsian idea that knowledge is a kind of standing in the space of reasons, a kind of status in the practice of the asking for and giving of reasons. But this correct insight is deformed by the "interiorization of the space of reasons," that is, the "withdrawal of it from the external world." One sees that no matter how flawless one's standing is in the space of reasons in regard to propositions about the external world, there is no assurance of truth. Thus, one seeks assurance by retreating inward, to propositions about how things appear. But once one so retreats, one never returns. The alternative to externalism is no less unattractive. In its full-blown form, it

denies that knowledge has anything to do with position in the space of reasons. Recognition of this weakness in pure externalism often leads externalists to affirm a hybrid conception of knowledge, according to which knowledge consists in both having the right sort of standing in the space of reasons and in being the recipient of the world's cooperation. McDowell attributes the following reasoning to hybrid theorists: since a subject can be perfectly justified in believing that p where $\langle p \rangle$ is false, in the instance in which one is perfectly justified in believing that p and $\langle p \rangle$ is true, one's knowledge is due in part to luck, in fact to the very same piece of luck that, by its absence, prevented knowledge in the former case. McDowell challenges this reasoning, asking how, if it were true, "reason [could] have the resources it would need in order to evaluate the reliability of belief-forming policies or habits?"

McDowell denies that flawless standing in the space of reasons could fall short of knowledge. Yet he agrees that there is no "luck-free zone." Why, then, is he himself not a hybrid theorist? The answer is that he maintains that when the world does us the favor of being the way it appears to be, this is part of our standing in the space of reasons, not something added to it. (It is interesting to compare this with the views defended by Susan Haack in Part IV and those of Ernest Sosa in Part V.)

Robert Brandom's stated aim in his contribution is to explain and develop the ideas expressed by McDowell. His chief concern is to describe the role of *social* practice in the determination of the existence and nature of the space of reasons, something he thinks is not made clear in McDowell's paper. The space of reasons is not a mind-independent Platonic object, but a creature of our social practice of adopting commitments and attributing commitments to others. The key link between truth and positive standing in the space of reasons is only revealed once we realize the role of attributing knowledge to others. In attributing knowledge we not only describe another's commitment, we take on that commitment ourselves. The paper concludes with the claim, defended at length in Brandom's *Making It Explicit*, that the social character of the space of reasons enables us to see how our various commitments could be about an objective world: "Our practice of comparing, assessing, and correcting different repertoires of commitments one with respect to another – those we attribute to others and those we undertake

ourselves – are what make them intelligible as *perspectives*, views of something, ways in which a

perspective-independent reality can *appear*.”

Further Reading

- Alston, William, “An Internalist Externalism,” in *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- , “Internalism and Externalism in Epistemology,” in *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- BonJour, L., *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985), esp. ch. 3.
- Feldman, Richard, “Reliability and Justification,” *The Monist* 68 (1985), 159–74.
- Foley, R., *Working Without a Net: A Study of Ego-centric Epistemology* (Oxford: Oxford University Press, 1993).
- Fumerton, Richard, “The Internalism/Externalism Controversy,” in James Tomberlin (ed.), *Philosophical Perspectives 2: Epistemology* (Atascadero, CA: Ridgeview Publishing Company, 1990), pp. 443–60.
- , *Metaepistemology and Skepticism* (Langham, MD: Rowman & Littlefield, 1995).
- Goldman, A., *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- , *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA, Massachusetts Institute of Technology Press, 1992).
- Greco, J., “Internalism and Epistemically Responsible Belief,” *Synthese* 85 (1990), pp. 245–77.
- Kornblith, H., “How Internal Can You Get?” *Synthese* 74 (1988), pp. 313–27.
- Moser, P., *Knowledge and Evidence* (Cambridge: Cambridge University Press, 1989).
- Nozick, Robert, *Philosophical Explanations* (Oxford: Oxford University Press, 1981).
- Plantinga, Alvin, *Warrant and Proper Function* (Oxford: Oxford University Press, 1993).
- , *Warrant: The Current Debate* (Oxford: Oxford University Press), esp. chs 1–3.
- Sosa, Ernest, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991).

What Is Justified Belief?

Alvin I. Goldman

The aim of this essay is to sketch a theory of justified belief. What I have in mind is an explanatory theory, one that explains in a general way why certain beliefs are counted as justified and others as unjustified. Unlike some traditional approaches, I do not try to prescribe standards for justification that differ from, or improve upon, our ordinary standards. I merely try to explicate the ordinary standards, which are, I believe, quite different from those of many classical, e.g., "Cartesian," accounts.

Many epistemologists have been interested in justification because of its presumed close relationship to knowledge. This relationship is intended to be preserved in the conception of justified belief presented here. In previous papers on knowledge,¹ I have denied that justification is necessary for knowing, but there I had in mind "Cartesian" accounts of justification. On the account of justified belief suggested here, it is necessary for knowing, and closely related to it.

The term "justified," I presume, is an evaluative term, a term of appraisal. Any correct definition or synonym of it would also feature evaluative terms. I assume that such definitions or synonyms might be given, but I am not interested in them. I want a set of *substantive* conditions that specify when a belief is justified. Compare the moral term "right." This might be defined in other ethical terms or phrases, a task appropriate to meta-ethics. The task of normative ethics, by contrast, is to state substantive conditions for the rightness of actions. Normative ethics tries to spe-

cify non-ethical conditions that determine when an action is right. A familiar example is act-utilitarianism, which says an action is right if and only if it produces, or would produce, at least as much net happiness as any alternative open to the agent. These necessary and sufficient conditions clearly involve no ethical notions. Analogously, I want a theory of justified belief to specify in non-epistemic terms when a belief is justified. This is not the only kind of theory of justifiedness one might seek, but it is one important kind of theory and the kind sought here.

In order to avoid epistemic terms in our theory, we must know which terms are epistemic. Obviously, an exhaustive list cannot be given, but here are some examples: "justified," "warranted," "has (good) grounds," "has reason (to believe)," "knows that," "sees that," "apprehends that," "is probable" (in an epistemic or inductive sense), "shows that," "establishes that," and "ascertains that." By contrast, here are some sample non-epistemic expressions: "believes that," "is true," "causes," "it is necessary that," "implies," "is deducible from," and "is probable" (either in the frequency sense or the propensity sense). In general, (purely) doxastic, metaphysical, modal, semantic, or syntactic expressions are not epistemic.

There is another constraint I wish to place on a theory of justified belief, in addition to the constraint that it be couched in non-epistemic language. Since I seek an explanatory theory, i.e., one that clarifies the underlying source of justificational status, it is not enough for a theory to state "correct" necessary and sufficient conditions. Its conditions must also be appropriately deep or

Originally published in G. S. Pappas (ed.), *Justification and Knowledge* (Dordrecht: D. Reidel, 1976), pp. 1-23;

revelatory. Suppose, for example, that the following sufficient condition of justified belief is offered: "If *S* senses redly at *t* and *S* believes at *t* that he is sensing redly, then *S*'s belief at *t* that he is sensing redly is justified." This is not the kind of principle I seek; for, even if it is correct, it leaves unexplained *why* a person who senses redly and believes that he does, believes this justifiably. Not every state is such that if one is in it and believes one is in it, this belief is justified. What is distinctive about the state of sensing redly, or "phenomenal" states in general? A theory of justified belief of the kind I seek must answer this question, and hence it must be couched at a suitably deep, general, or abstract level.

A few introductory words about my *explicandum* are appropriate at this juncture. It is often assumed that whenever a person has a justified belief, he knows that it is justified and knows what the justification is. It is further assumed that the person can state or explain what his justification is. On this view, a justification is an argument, defense, or set of reasons that can be given in support of a belief. Thus, one studies the nature of justified belief by considering what a person might say if asked to defend, or justify, his belief. I make none of these sorts of assumptions here. I leave it an open question whether, when a belief is justified, the believer *knows* it is justified. I also leave it an open question whether, when a belief is justified, the believer can *state* or *give* a justification for it. I do not even assume that when a belief is justified there is something "possessed" by the believer which can be called a "justification." I do assume that a justified belief gets its status of being justified from some processes or properties that make it justified. In short, there must be some justification-conferring processes or properties. But this does not imply that there must be an argument, or reason, or anything else, "possessed" at the time of belief by the believer.

I

A theory of justified belief will be a set of principles that specify truth-conditions for the schema [*S*'s belief in *p* at time *t* is justified], i.e., conditions for the satisfaction of this schema in all possible cases. It will be convenient to formulate candidate theories in a recursive or inductive format, which would include (A) one or more base clauses, (B) a set of recursive clauses (possibly

null), and (C) a closure clause. In such a format, it is permissible for the predicate "is a justified belief" to appear in recursive clauses. But neither this predicate, nor any other epistemic predicate, may appear in (the antecedent of) any base clause.²

Before turning to my own theory, I want to survey some other possible approaches to justified belief. Identification of problems associated with other attempts will provide some motivation for the theory I shall offer. Obviously, I cannot examine all, or even very many, alternative attempts. But a few sample attempts will be instructive.

Let us concentrate on the attempt to formulate one or more adequate base-clause principles.³ Here is a classical candidate:

- (1) If *S* believes *p* at *t*, and *p* is indubitable for *S* (at *t*), then *S*'s belief in *p* at *t* is justified.

To evaluate this principle, we need to know what "indubitable" means. It can be understood in at least two ways. First, "*p* is indubitable for *S*" might mean: "*S* has no grounds for doubting *p*." Since "ground" is an epistemic term, however, principle (1) would be inadmissible in this reading, for epistemic terms may not legitimately appear in the antecedent of a base clause. A second interpretation would avoid this difficulty. One might interpret "*p* is indubitable for *S*" psychologically, i.e., as meaning "*S* is psychologically incapable of doubting *p*." This would make principle (1) admissible, but would it be correct? Surely not. A religious fanatic may be psychologically incapable of doubting the tenets of his faith, but that doesn't make his belief in them justified. Similarly, during the Watergate affair, someone may have been so blinded by the aura of the presidency that even after the most damaging evidence against Nixon had emerged he was still incapable of doubting Nixon's veracity. It doesn't follow that his belief in Nixon's veracity was justified.

A second candidate base-clause principle is this:

- (2) If *S* believes *p* at *t* and *p* is self-evident, then *S*'s belief in *p* at *t* is justified.

To evaluate this principle, we again need an interpretation of its crucial term, in this case "self-evident." On one standard reading, "evident" is a synonym for "justified." "Self-evident" would therefore mean something like "directly justified," "intuitively justified," or "non-derivatively justified." On this reading "self-evident" is an epistemic phrase, and principle (2) would be disqualified as a base-clause principle.

However, there are other possible readings of “ p is self-evident” on which it isn’t an epistemic phrase. One such reading is: “It is impossible to understand p without believing it.”⁴ According to this interpretation, trivial analytic and logical truths might turn out to be self-evident. Hence, any belief in such a truth would be a justified belief, according to (2).

What does “it is impossible to understand p without believing it” mean? Does it mean “humanly impossible”? That reading would probably make (2) an unacceptable principle. There may well be propositions which humans have an innate and irrepressible disposition to believe, e.g., “Some events have causes.” But it seems unlikely that people’s inability to refrain from believing such a proposition makes every belief in it justified.

Should we then understand “impossible” to mean “impossible in principle,” or “logically impossible”? If that is the reading given, I suspect that (2) is a vacuous principle. I doubt that even trivial logical or analytic truths will satisfy this definition of “self-evident.” Any proposition, we may assume, has two or more components that are somehow organized or juxtaposed. To understand the proposition one must “grasp” the components and their juxtaposition. Now in the case of *complex* logical truths, there are (human) psychological operations that suffice to grasp the components and their juxtaposition but do not suffice to produce a belief that the proposition is true. But can’t we at least *conceive* of an analogous set of psychological operations even for simple logical truths, operations which perhaps are not in the repertoire of human cognizers but which might be in the repertoire of some conceivable beings? That is, can’t we conceive of psychological operations that would suffice to grasp the components and component-juxtaposition of these simple propositions but do not suffice to produce *belief* in the propositions? I think we can conceive of such operations. Hence, for any proposition you choose, it will be possible for it to be understood without being believed.

Finally, even if we set these two objections aside, we must note that self-evidence can at best confer justificational status on relatively few beliefs, and the only plausible group are beliefs in necessary truths. Thus, other base-clause principles will be needed to explain the justificational status of beliefs in contingent propositions.

The notion of a base-clause principle is naturally associated with the idea of “direct” justified-

ness, and in the realm of contingent propositions first-person-current-mental-state propositions have often been assigned this role. In Chisholm’s terminology, this conception is expressed by the notion of a “self-presenting” state or proposition. The sentence “I am thinking,” for example, expresses a self-presenting proposition. (At least I shall call this sort of content a “proposition,” though it only has a truth value given some assignment of a subject who utters or entertains the content and a time of entertaining.) When such a proposition is true for person S at time t , S is justified in believing it at t : in Chisholm’s terminology, the proposition is “evident” for S at t . This suggests the following base-clause principle.

- (3) If p is a self-presenting proposition, and p is true for S at t , and S believes p at t , then S ’s belief in p at t is justified.

What, exactly, does “self-presenting” mean? In the second edition of *Theory of Knowledge*, Chisholm offers this definition: “ h is self-presenting for S at $t = df.$ h is true at t ; and necessarily, if h is true at t , then h is evident for S at t .”⁵ Unfortunately, since “evident” is an epistemic term, “self-presenting” also becomes an epistemic term on this definition, thereby disqualifying (3) as a legitimate base clause. Some other definition of self-presentingness must be offered if (3) is to be a suitable base-clause principle.

Another definition of self-presentation readily comes to mind. “Self-presentation” is an approximate synonym of “self-intimation,” and a proposition may be said to be self-intimating if and only if whenever it is true of a person that person believes it. More precisely, we may give the following definition:

- (SP) Proposition p is self-presenting if and only if: necessarily, for any S and any t , if p is true for S at t , then S believes p at t .

On this definition, “self-presenting” is clearly not an epistemic predicate, so (3) would be an admissible principle. Moreover, there is initial plausibility in the suggestion that it is *this* feature of first-person-current-mental-state propositions – viz., their truth guarantees their being believed – that makes beliefs in them justified.

Employing this definition of self-presentation, is principle (3) correct? This cannot be decided until we define self-presentation more precisely. Since the operator “necessarily” can be read in different ways, there are different forms of

self-presentation and correspondingly different versions of principle (3). Let us focus on two of these readings: a “*nomological*” reading and a “*logical*” reading. Consider first the nomological reading. On this definition a proposition is self-presenting just in case it is nomologically necessary that if p is true for S at t , then S believes p at t .⁶

Is the nomological version of principle (3) – call it “(3_N)” – correct? Not at all. We can imagine cases in which the antecedent of (3_N) is satisfied, but we would not say that the belief is justified. Suppose, for example, that p is the proposition expressed by the sentence “I am in brain-state B ,” where “ B ” is shorthand for a certain highly specific neural state description. Further suppose it is a nomological truth that anyone in brain-state B will ipso facto believe he is in brain-state B . In other words, imagine that an occurrent belief with the content “I am in brain-state B ” is realized whenever one is in brain-state B .⁷ According to (3_N), any such belief is justified. But that is clearly false. We can readily imagine circumstances in which a person goes into brain-state B and therefore has the belief in question, though this belief is by no means justified. For example, we can imagine that a brain-surgeon operating on S artificially induced brain-state B . This results, phenomenologically, in S 's suddenly believing – out of the blue – that he is in brain-state B , without any relevant antecedent beliefs. We would hardly say, in such a case, that S 's belief that he is in brain-state B is justified.

Let us turn next to the logical version of (3) – call it “(3_L)” – in which a proposition is defined as self-presenting just in case it is logically necessary that if p is true for S at t , then S believes p at t . This stronger version of principle (3) might seem more promising. In fact, however, it is no more successful than (3_N). Let p be the proposition “I am awake” and assume that it is logically necessary that if this proposition is true for some person S and time t , then S believes p at t . This assumption is consistent with the further assumption that S frequently believes p when it is false, e.g., when he is dreaming. Under these circumstances, we would hardly accept the contention that S 's belief in this proposition is always justified. Nor should we accept the contention that the belief is justified when it is true. The truth of the proposition logically guarantees that the belief is held, but why should it guarantee that the belief is justified?

The foregoing criticism suggests that we have things backwards. The idea of self-presentation is

that truth guarantees belief. This fails to confer justification because it is compatible with there being belief without truth. So what seems necessary – or at least sufficient – for justification is that belief should guarantee truth. Such a notion has usually gone under the label of “*infallibility*” or “*incorrigibility*.” It may be defined as follows:

(INC) Proposition p is incorrigible if and only if: necessarily, for any S and any t , if S believes p at t , then p is true for S at t .

Using the notion of incorrigibility, we may propose principle (4).

(4) If p is an incorrigible proposition, and S believes p at t , then S 's belief in p at t is justified.

As was true of self-presentation, there are different varieties of incorrigibility, corresponding to different interpretations of “necessarily.” Accordingly, we have different versions of principle (4). Once again, let us concentrate on a nomological and a logical version, (4_N) and (4_L) respectively.

We can easily construct a counterexample to (4_N) along the lines of the belief-state/brain-state counterexample that refuted (3_N). Suppose it is nomologically necessary that if anyone believes he is in brain-state B then it is true that he is in brain-state B , for the only way this belief-state is realized is through brain-state B itself. It follows that “I am in brain-state B ” is a nomologically incorrigible proposition. Therefore, according to (4_N), whenever anyone believes this proposition at any time, that belief is justified. But we may again construct a brain-surgeon example in which someone comes to have such a belief but the belief isn't justified.

Apart from this counterexample, the general point is this. Why should the fact that S 's believing p guarantees the truth of p imply that S 's belief is justified? The nature of the guarantee might be wholly fortuitous, as the belief-state/brain-state example is intended to illustrate. To appreciate the point, consider the following related possibility. A person's mental structure might be such that whenever he believes that p will be true (of him) a split second later, then p is true (of him) a split second later. This is because, we may suppose, his believing it brings it about. But surely we would not be compelled in such a circumstance to say that a belief of this sort is justified. So why should the fact that S 's believing p guarantees the truth of p precisely at the time of belief imply that

the belief is justified? There is no intuitive plausibility in this supposition.

The notion of *logical* incorrigibility has a more honored place in the history of conceptions of justification. But even principle (4_L), I believe, suffers from defects similar to those of (4_N). The mere fact that belief in *p* logically guarantees its truth does not confer justificational status on such a belief.

The first difficulty with (4_L) arises from logical or mathematical truths. Any true proposition of logic or mathematics is logically necessary. Hence, any such proposition *p* is logically incorrigible, since it is logically necessary that, for any *S* and any *t*, if *S* believes *p* at *t* then *p* is true (for *S* at *t*). Now assume that Nelson believes a certain very complex mathematical truth at time *t*. Since such a proposition is logically incorrigible, (4_L) implies that Nelson's belief in this truth at *t* is justified. But we may easily suppose that this belief of Nelson is not at all the result of proper mathematical reasoning, or even the result of appeal to trustworthy authority. Perhaps Nelson believes this complex truth because of utterly confused reasoning, or because of hasty and ill-founded conjecture. Then his belief is not justified, contrary to what (4_L) implies.

The case of logical or mathematical truths is admittedly peculiar, since the truth of these propositions is assured independently of any beliefs. It might seem, therefore, that we can better capture the idea of "belief logically guaranteeing truth" in cases where the propositions in question are *contingent*. With this in mind, we might restrict (4_L) to *contingent* incorrigible propositions. Even this amendment cannot save (4_L), however, since there are counterexamples to it involving purely contingent propositions.

Suppose that Humperdink has been studying logic – or, rather, pseudo-logic – from Elmer Fraud, whom Humperdink has no reason to trust as a logician. Fraud has enunciated the principle that any disjunctive proposition consisting of at least 40 distinct disjuncts is very probably true. Humperdink now encounters the proposition *p*, a contingent proposition with 40 disjuncts, the 7th disjunct being "I exist." Although Humperdink grasps the proposition fully, he doesn't notice that it is entailed by "I exist." Rather, he is struck by the fact that it falls under the disjunction rule Fraud has enunciated (a rule I assume Humperdink is not *justified* in believing). Bearing this in mind, Humperdink forms a belief in *p*. Now notice

that *p* is logically incorrigible. It is logically necessary that if anyone believes *p*, then *p* is true (of him at that time). This simply follows from the fact that, first, a person's believing anything entails that he exists, and second, "I exist" entails *p*. Since *p* is logically incorrigible, principle (4_L) implies that Humperdink's belief in *p* is justified. But surely, given our example, that conclusion is false. Humperdink's belief in *p* is not at all justified.

One thing that goes wrong in this example is that while Humperdink's belief in *p* logically implies its truth, Humperdink doesn't *recognize* that his believing it implies its truth. This might move a theorist to revise (4_L) by adding the requirement that *S* "recognize" that *p* is logically incorrigible. But this, of course, won't do. The term "recognize" is obviously an epistemic term, so the suggested revision of (4_L) would result in an inadmissible base clause.

II

Let us try to diagnose what has gone wrong with these attempts to produce an acceptable base-clause principle. Notice that each of the foregoing attempts confers the status of "justified" on a belief without restriction on *why* the belief is held, i.e., on what *causally initiates* the belief or *causally sustains* it. The logical versions of principles (3) and (4), for example, clearly place no restriction on causes of belief. The same is true of the nomological versions of (3) and (4), since nomological requirements can be satisfied by simultaneity or cross-sectional laws, as illustrated by our brain-state/belief-state examples. I suggest that the absence of causal requirements accounts for the failure of the foregoing principles. Many of our counterexamples are ones in which the belief is caused in some strange or unacceptable way, e.g., by the accidental movement of a brain-surgeon's hand, by reliance on an illicit, pseudo-logical principle, or by the blinding aura of the presidency. In general, a strategy for defeating a noncausal principle of justifiedness is to find a case in which the principle's antecedent is satisfied but the belief is caused by some faulty belief-forming process. The faultiness of the belief-forming process will incline us, intuitively, to regard the belief as unjustified. Thus, correct principles of justified belief must be principles that make causal requirements, where "cause" is construed broadly to include sustainers

as well as initiators of belief (i.e., processes that determine, or help to overdetermine, a belief's continuing to be held).⁸

The need for causal requirements is not restricted to base-clause principles. Recursive principles will also need a causal component. One might initially suppose that the following is a good recursive principle: "If S justifiably believes q at t , and q entails p , and S believes p at t , then S's belief in p at t is justified." But this principle is unacceptable. S's belief in p doesn't receive justificational status simply from the fact that p is entailed by q and S justifiably believes q . If what causes S to believe p at t is entirely different, S's belief in p may well not be justified. Nor can the situation be remedied by adding to the antecedent the condition that S justifiably believes that q entails p . Even if he believes this, and believes q as well, he might not put these beliefs together. He might believe p as a result of some other wholly extraneous considerations. So once again, conditions that fail to require appropriate causes of a belief don't guarantee justifiedness.

Granted that principles of justified belief must make reference to causes of belief, what kinds of causes confer justifiedness? We can gain insight into this problem by reviewing some faulty processes of belief-formation, i.e., processes whose belief-outputs would be classed as unjustified. Here are some examples: confused reasoning, wishful thinking, reliance on emotional attachment, mere hunch or guesswork, and hasty generalization. What do these faulty processes have in common? They share the feature of *unreliability*: they tend to produce *error* a large proportion of the time. By contrast, which species of belief-forming (or belief-sustaining) processes are intuitively justification-conferring? They include standard perceptual processes, remembering, good reasoning, and introspection. What these processes seem to have in common is *reliability*: the beliefs they produce are generally true. My positive proposal, then, is this. The justificational status of a belief is a function of the reliability of the process or processes that cause it, where (as a first approximation) reliability consists in the tendency of a process to produce beliefs that are true rather than false.

To test this thesis further, notice that justifiedness is not a purely categorical concept, although I treat it here as categorical in the interest of simplicity. We can and do regard certain beliefs as more justified than others. Furthermore, our intuitions

of comparative justifiedness go along with our beliefs about the comparative reliability of the belief-causing processes.

Consider perceptual beliefs. Suppose Jones believes he has just seen a mountain-goat. Our assessment of the belief's justifiedness is determined by whether he caught a brief glimpse of the creature at a great distance, or whether he had a good look at the thing only 30 yards away. His belief in the latter sort of case is (*ceteris paribus*) more justified than in the former sort of case. And, if his belief is true, we are more prepared to say he *knows* in the latter case than in the former. The difference between the two cases seems to be this. Visual beliefs formed from brief and hasty scanning, or where the perceptual object is a long distance off, tend to be wrong more often than visual beliefs formed from detailed and leisurely scanning, or where the object is in reasonable proximity. In short, the visual processes in the former category are less reliable than those in the latter category. A similar point holds for memory beliefs. A belief that results from a hazy and indistinct memory impression is counted as less justified than a belief that arises from a distinct memory impression, and our inclination to classify those beliefs as "*knowledge*" varies in the same way. Again, the reason is associated with the comparative reliability of the processes. Hazy and indistinct memory impressions are generally less reliable indicators of what actually happened, so beliefs formed from such impressions are less likely to be true than beliefs formed from distinct impressions. Further, consider beliefs based on inference from observed samples. A belief about a population that is based on random sampling, or on instances that exhibit great variety, is intuitively more justified than a belief based on biased sampling, or on instances from a narrow sector of the population. Again, the degree of justifiedness seems to be a function of reliability. Inferences based on random or varied samples will tend to produce less error or inaccuracy than inferences based on non-random or non-varied samples.

Returning to a categorical concept of justifiedness, we might ask just *how* reliable a belief-forming process must be in order that its resultant beliefs be justified. A precise answer to this question should not be expected. Our conception of justification is *vague* in this respect. It does seem clear, however, that *perfect* reliability isn't required. Belief-forming processes that *sometimes*

produce error still confer justification. It follows that there can be justified beliefs that are false.

I have characterized justification-conferring processes as ones that have a "tendency" to produce beliefs that are true rather than false. The term "tendency" could refer either to *actual* long-run frequency, or to a "propensity," i.e., outcomes that would occur in merely *possible* realizations of the process. Which of these is intended? Unfortunately, I think our ordinary conception of justifiedness is vague on this dimension too. For the most part, we simply assume that the "observed" frequency of truth versus error would be approximately replicated in the actual long-run, and also in relevant counterfactual situations, i.e., ones that are highly "realistic" or conform closely to the circumstances of the actual world. Since we ordinarily assume these frequencies to be roughly the same, we make no concerted effort to distinguish them. Since the purpose of my present theorizing is to capture our ordinary conception of justifiedness, and since our ordinary conception is vague on this matter, it is appropriate to leave the theory vague in the same respect.

We need to say more about the notion of a belief-forming "*process*." Let us mean by a "process" a *functional operation* or procedure, i.e., something that generates a *mapping* from certain states – "inputs" – into other states – "outputs." The outputs in the present case are states of believing this or that proposition at a given moment. On this interpretation, a process is a *type* as opposed to a *token*. This is fully appropriate, since it is only types that have statistical properties such as producing truth 80 per cent of the time; and it is precisely such statistical properties that determine the reliability of a process. Of course, we also want to speak of a process as *causing* a belief, and it looks as if types are incapable of being causes. But when we say that a belief is caused by a given process, understood as a functional procedure, we may interpret this to mean that it is caused by the particular *inputs* to the process (and by the intervening events "through which" the functional procedure carries the inputs into the output) on the occasion in question.

What are some examples of belief-forming "processes" construed as functional operations? One example is reasoning processes, where the inputs include antecedent beliefs and entertained hypotheses. Another example is functional procedures whose inputs include desires, hopes, or emo-

tional states of various sorts (together with antecedent beliefs). A third example is a memory process, which takes as input beliefs or experiences at an earlier time and generates as output beliefs at a later time. For example, a memory process might take as input a belief at t_1 that Lincoln was born in 1809 and generate as output a belief at t_n that Lincoln was born in 1809. A fourth example is perceptual processes. Here it isn't clear whether inputs should include states of the environment, such as the distance of the stimulus from the cognizer, or only events within or on the surface of the organism, e.g., receptor stimulations. I shall return to this point in a moment.

A critical problem concerning our analysis is the degree of generality of the process-types in question. Input-output relations can be specified very broadly or very narrowly, and the degree of generality will partly determine the degree of reliability. A process-type might be selected so narrowly that only one instance of it ever occurs, and hence the type is either completely reliable or completely unreliable. (This assumes that reliability is a function of *actual* frequency only.) If such narrow process-types were selected, beliefs that are intuitively unjustified might be said to result from perfectly reliable processes, and beliefs that are intuitively justified might be said to result from perfectly unreliable processes.

It is clear that our ordinary thought about process-types slices them broadly, but I cannot at present give a precise explication of our intuitive principles. One plausible suggestion, though, is that the relevant processes are *content-neutral*. It might be argued, for example, that the process of *inferring p whenever the Pope asserts p* could pose problems for our theory. If the Pope is infallible, this process will be perfectly reliable; yet we would not regard the belief-outputs of this process as justified. The content-neutral restriction would avert this difficulty. If relevant processes are required to admit as input beliefs (or other states) with *any* content, the aforementioned process will not count, for its input beliefs have a restricted propositional content, viz., "*the Pope asserts p*."

In addition to the problem of "generality" or "abstractness" there is the previously mentioned problem of the "*extent*" of belief-forming processes. Clearly, the causal ancestry of beliefs often includes events outside the organism. Are such events to be included among the "inputs" of belief-forming processes? Or should we restrict the extent of belief-forming processes to

“cognitive” events, i.e., events within the organism’s nervous system? I shall choose the latter course, though with some hesitation. My general grounds for this decision are roughly as follows. Justifiedness seems to be a function of how a cognizer deals with his environmental input, i.e., with the goodness or badness of the operations that register and transform the stimulation that reaches him. (“Deal with,” of course, does not mean *purposeful* action, nor is it restricted to *conscious* activity.) A justified belief is, roughly speaking, one that results from cognitive operations that are, generally speaking, good or successful. But “cognitive” operations are most plausibly construed as operations of the cognitive faculties, i.e., “information-processing” equipment *internal* to the organism.

With these points in mind, we may now advance the following base-clause principle for justified belief.

- (5) If S’s believing *p* at *t* results from a reliable cognitive belief-forming process (or set of processes), then S’s belief in *p* at *t* is justified.

Since “reliable belief-forming process” has been defined in terms of such notions as belief, truth, statistical frequency, and the like, it is not an epistemic term. Hence, (5) is an admissible base clause.

It might seem as if (5) promises to be not only a successful base clause, but the only principle needed whatever, apart from a closure clause. In other words, it might seem as if it is a necessary as well as a sufficient condition of justifiedness that a belief be produced by reliable cognitive belief-forming processes. But this is not quite correct, given our provisional definition of “reliability.”

Our provisional definition implies that a reasoning process is reliable only if it generally produces beliefs that are true, and similarly, that a memory process is reliable only if it generally yields beliefs that are true. But these requirements are too strong. A reasoning procedure cannot be expected to produce true belief if it is applied to false premises. And memory cannot be expected to yield a true belief if the original belief it attempts to retain is false. What we need for reasoning and memory, then, is a notion of “*conditional reliability*.” A process is conditionally reliable when a sufficient proportion of its output-beliefs are true *given that its input-beliefs are true*.

With this point in mind, let us distinguish *belief-dependent* and *belief-independent* cognitive pro-

cesses. The former are processes *some* of whose inputs are belief-states.⁹ The latter are processes *none* of whose inputs are belief-states. We may then replace principle (5) with the following two principles, the first a base-clause principle and the second a recursive-clause principle.

- (6A) If S’s belief in *p* at *t* results (“immediately”) from a belief-independent process that is (unconditionally) reliable, then S’s belief in *p* at *t* is justified.
- (6B) If S’s belief in *p* at *t* results (“immediately”) from a belief-dependent process that is (at least) conditionally reliable, and if the beliefs (if any) on which this process operates in producing S’s belief in *p* at *t* are themselves justified, then S’s belief in *p* at *t* is justified.¹⁰

If we add to (6A) and (6B) the standard closure clause, we have a complete theory of justified belief. The theory says, in effect, that a belief is justified if and only if it is “*well-formed*,” i.e., it has an ancestry of reliable and/or conditionally reliable cognitive operations. (Since a dated belief may be over-determined, it may have a number of distinct ancestral trees. These need not all be full of reliable or conditionally reliable processes. But at least one ancestral tree must have reliable or conditionally reliable processes throughout.)

The theory of justified belief proposed here, then, is an *Historical* or *Genetic* theory. It contrasts with the dominant approach to justified belief, an approach that generates what we may call (borrowing a phrase from Robert Nozick) “*Current Time-Slice*” theories. A Current Time-Slice theory makes the justificational status of a belief wholly a function of what is true of the cognizer *at the time* of belief. An Historical theory makes the justificational status of a belief depend on its prior history. Since my Historical theory emphasizes the reliability of the belief-generating processes, it may be called “*Historical Reliabilism*.”

The most obvious examples of Current Time-Slice theories are “*Cartesian*” Foundationalist theories, which trace all justificational status (at least of contingent propositions) to current mental states. The usual varieties of Coherence theories, however, are equally Current Time-Slice views, since they too make the justificational status of a belief wholly a function of *current* states of affairs. For Coherence theories, however, these current states include all other beliefs of the cognizer, which would not be considered relevant by

Cartesian Foundationalism. Have there been other Historical theories of justified belief? Among contemporary writers, Quine and Popper have Historical epistemologies, though the notion of "justification" is not their avowed *explicandum*. Among historical writers, it might seem that Locke and Hume had Genetic theories of sorts. But I think that their Genetic theories were only theories of ideas, not of knowledge or justification. Plato's theory of recollection, however, is a good example of a Genetic theory of knowing.¹¹ And it might be argued that Hegel and Dewey had Genetic epistemologies (if Hegel can be said to have had a clear epistemology at all).

The theory articulated by (6A) and (6B) might be viewed as a kind of "Foundationalism" because of its recursive structure. I have no objection to this label, as long as one keeps in mind how different this "diachronic" form of Foundationalism is from Cartesian, or other "synchronic" varieties of Foundationalism.

Current Time-Slice theories characteristically assume that the justificational status of a belief is something which the cognizer is able to know or determine at the time of belief. This is made explicit, for example, by Chisholm.¹² The Historical theory I endorse makes no such assumption. There are many facts about a cognizer to which he lacks "privileged access," and I regard the justificational status of his beliefs as one of those things. This is not to say that a cognizer is necessarily ignorant, at any given moment, of the justificational status of his current beliefs. It is only to deny that he necessarily has, or can get, knowledge or true belief about this status. Just as a person can know without knowing that he knows, so he can have justified belief without knowing that it is justified (or believing justifiably that it is justified).

A characteristic case in which a belief is justified though the cognizer doesn't know that it's justified is where the original evidence for the belief has long since been forgotten. If the original evidence was compelling, the cognizer's original belief may have been justified, and this justificational status may have been preserved through memory. But since the cognizer no longer remembers how or why he came to believe, he may not know that the belief is justified. If asked now to justify his belief, he may be at a loss. Still, the belief is justified, though the cognizer can't demonstrate or establish this.

The Historical theory of justified belief I advocate is connected in spirit with the causal theory of

knowing I have presented elsewhere.¹³ I had this in mind when I remarked near the outset of the essay that my theory of justified belief makes justifiedness come out closely related to knowledge. Justified beliefs, like pieces of knowledge, have appropriate histories; but they may fail to be knowledge either because they are false or because they founder on some other requirement for knowing of the kind discussed in the post-Gettier knowledge-trade.

There is a variant of the Historical conception of justified belief that is worth mentioning in this context. It may be introduced as follows. Suppose S has a set *B* of beliefs at time t_0 , and some of these beliefs are *unjustified*. Between t_0 and t_1 he reasons from the entire set *B* to the conclusion *p*, which he then accepts at t_1 . The reasoning procedure he uses is a very sound one, i.e., one that is conditionally reliable. There is a sense or respect in which we are tempted to say that S's belief in *p* at t_1 is "justified." At any rate, it is tempting to say that the *person* is justified in believing *p* at t . Relative to his antecedent cognitive state, he did as well as could be expected: the *transition* from his cognitive state at t_0 to his cognitive state at t_1 was entirely sound. Although we may acknowledge this brand of justifiedness – it might be called "*Terminal-Phase Reliabilism*" – it is not a kind of justifiedness so closely related to knowing. For a person to know proposition *p*, it is not enough that the *final phase* of the process that leads to his belief in *p* be sound. It is also necessary that some entire history of the process be sound (i.e., reliable or conditionally reliable).

Let us return now to the Historical theory. In the next section, I shall adduce reasons for strengthening it a bit. Before looking at these reasons, however, I wish to review two quite different objections to the theory.

First, a critic might argue that *some* justified beliefs do not derive their justificational status from their causal ancestry. In particular, it might be argued that beliefs about one's current phenomenal states and intuitive beliefs about elementary logical or conceptual relationships do not derive their justificational status in this way. I am not persuaded by either of these examples. Introspection, I believe, should be regarded as a form of retrospection. Thus, a justified belief that I am "now" in pain gets its justificational status from a relevant, though brief, causal history.¹⁴ The apprehension of logical or conceptual relationships is also a cognitive process that occupies time. The

psychological process of “seeing” or “intuiting” a simple logical truth is very fast, and we cannot introspectively dissect it into constituent parts. Nonetheless, there are mental operations going on, just as there are mental operations that occur in *idiots savants*, who are unable to report the computational processes they in fact employ.

A second objection to Historical Reliabilism focuses on the reliability element rather than the causal or historical element. Since the theory is intended to cover all possible cases, it seems to imply that for any cognitive process *C*, if *C* is reliable in possible world *W*, then any belief in *W* that results from *C* is justified. But doesn't this permit easy counterexamples? Surely we can imagine a possible world in which wishful thinking is reliable. We can imagine a possible world where a benevolent demon so arranges things that beliefs formed by wishful thinking usually come true. This would make wishful thinking a reliable process in that possible world, but surely we don't want to regard beliefs that result from wishful thinking as justified.

There are several possible ways to respond to this case, and I am unsure which response is best, partly because my own intuitions (and those of other people I have consulted) are not entirely clear. One possibility is to say that in the possible world imagined, beliefs that result from wishful thinking *are* justified. In other words, we reject the claim that wishful thinking could never, intuitively, confer justifiedness.¹⁵

However, for those who feel that wishful thinking couldn't confer justifiedness even in the world imagined, there are two ways out. First, it may be suggested that the proper criterion of justifiedness is the propensity of a process to generate beliefs that are true *in a non-manipulated environment*, i.e., an environment in which there is no purposeful arrangement of the world either to accord or conflict with the beliefs that are formed. In other words, the suitability of a belief-forming process is only a function of its success in “*natural*” situations, not situations of the sort involving benevolent or malevolent demons or any other such manipulative creatures. If we reformulate the theory to include this qualification, the counterexample in question will be averted.

Alternatively, we may reformulate our theory, or reinterpret it, as follows. Instead of construing the theory as saying that a belief in possible world *W* is justified if and only if it results from a cognitive process that is reliable in *W*, we may

construe it as saying that a belief in possible world *W* is justified if and only if it results from a cognitive process that is reliable in *our world*. In short, our conception of justifiedness is derived as follows. We note certain cognitive processes in the actual world, and form beliefs about which of these are reliable. The ones we believe to be reliable are then regarded as justification-conferring processes. In reflecting on hypothetical beliefs, we deem them justified if and only if they result from processes already picked out as justification-conferring, or processes very similar to those. Since wishful thinking is not among these processes, a belief formed in a possible world *W* by wishful thinking would not be deemed justified, even if wishful thinking is reliable *in W*. I am not sure that this is a correct reconstruction of our intuitive conceptual scheme, but it would accommodate the benevolent demon case, at least if the proper thing to say in that case in that the wishful-thinking-caused beliefs are unjustified.

Even if we adopt this strategy, however, a problem still remains. Suppose that wishful thinking turns out to be reliable *in the actual world!*¹⁶ This might be because, unbeknownst to us at present, there is a benevolent demon who, lazy until now, will shortly start arranging things so that our wishes come true. The long-run performance of wishful thinking will be very good, and hence even the new construal of the theory will imply that beliefs resulting from wishful thinking (in *our world*) are justified. Yet this surely contravenes our intuitive judgment on the matter.

Perhaps the moral of the case is that the standard format of a “conceptual analysis” has its shortcomings. Let me depart from that format and try to give a better rendering of our aim and the theory that tries to achieve that aim. What we really want is an *explanation* of why we count, or would count, certain beliefs as justified and others as unjustified. Such an explanation must refer to our *beliefs* about reliability, not to the actual *facts*. The reason we *count* beliefs as justified is that they are formed by what we *believe* to be reliable belief-forming processes. Our beliefs about which belief-forming processes are reliable may be erroneous, but that does not affect the adequacy of the explanation. Since we *believe* that wishful thinking is an unreliable belief-forming process, we regard beliefs formed by wishful thinking as unjustified. What matters, then, is what we *believe* about wishful thinking, not what is *true* (in the long run) about wishful thinking. I am not sure how to express this

point in the standard format of conceptual analysis, but it identifies an important point in understanding our theory.

III

Let us return, however, to the standard format of conceptual analysis, and let us consider a new objection that will require some revisions in the theory advanced until now. According to our theory, a belief is justified in case it is caused by a process that is in fact reliable, or by one we generally believe to be reliable. But suppose that although one of S's beliefs satisfies this condition, S has no reason to believe that it does. Worse yet, suppose S has reason to believe that his belief is caused by an *unreliable* process (although *in fact* its causal ancestry is fully reliable). Wouldn't we deny in such circumstances that S's belief is justified? This seems to show that our analysis, as presently formulated, is mistaken.

Suppose that Jones is told on fully reliable authority that a certain class of his memory beliefs are almost all mistaken. His parents fabricate a wholly false story that Jones suffered from amnesia when he was seven but later developed *pseudo*-memories of that period. Though Jones listens to what his parents say and has excellent reason to trust them, he persists in believing the ostensible memories from his seven-year-old past. Are these memory beliefs justified? Intuitively, they are not justified. But since these beliefs result from genuine memory and original perceptions, which are adequately reliable processes, our theory says that these beliefs are justified.

Can the theory be revised to meet this difficulty? One natural suggestion is that the actual reliability of a belief's ancestry is not enough for justifiedness; in addition, the cognizer must be *justified in believing* that the ancestry of his belief is reliable. Thus one might think of replacing (6A), for example, with (7). (For simplicity, I neglect some of the details of the earlier analysis.)

- (7) If S's belief in p at t is caused by a reliable cognitive process, and S justifiably believes at t that his p -belief is so caused, then S's belief in p at t is justified.

It is evident, however, that (7) will not do as a base clause, for it contains the epistemic term "justifiably" in its antecedent.

A slightly weaker revision, without this problematic feature, might next be suggested, viz.,

- (8) If S's belief in p at t is caused by a reliable cognitive process, and S believes at t that his p -belief is so caused, then S's belief in p at t is justified.

But this won't do the job. Suppose that Jones believes that his memory beliefs are reliably caused despite all the (trustworthy) contrary testimony of his parents. Principle (8) would be satisfied, yet we wouldn't say that these beliefs are justified.

Next, we might try (9), which is stronger than (8) and, unlike (7), formally admissible as a base clause.

- (9) If S's belief in p at t is caused by a reliable cognitive process, and S believes at t that his p -belief is so caused, and this meta-belief is caused by a reliable cognitive process, then S's belief in p at t is justified.

A first objection to (9) is that it wrongly precludes unreflective creatures – creatures like animals or young children, who have no beliefs about the genesis of their beliefs – from having justified beliefs. If one shares my view that justified belief is, at least roughly, *well-formed* belief, surely animals and young children can have justified beliefs.

A second problem with (9) concerns its underlying rationale. Since (9) is proposed as a substitute for (6A), it is implied that the reliability of a belief's own cognitive ancestry does not make it justified. But, the suggestion seems to be, the reliability of a *meta-belief's* ancestry confers justifiedness on the first-order belief. Why should that be so? Perhaps one is attracted by the idea of a "trickle-down" effect: if an $n + 1$ -level belief is justified, its justification trickles down to an n -level belief. But even if the trickle-down theory is correct, it doesn't help here. There is no assurance from the satisfaction of (9)'s antecedent that the meta-belief itself is *justified*.

To obtain a better revision of our theory, let us re-examine the Jones case. Jones has strong evidence against certain propositions concerning his past. He doesn't *use* this evidence, but if he *were* to use it properly, he would stop believing these propositions. Now the proper use of evidence would be an instance of a (conditionally) reliable process. So what we can say about Jones is that he *fails* to use a certain (conditionally) reliable process that he could and should have used. Admittedly, had he used this process, he would

have "worsened" his doxastic states: he would have replaced some true beliefs with suspension of judgment. Still, he couldn't have known this in the case in question. So he failed to do something which, epistemically, he should have done. This diagnosis suggests a fundamental change in our theory. The justificational status of a belief is not only a function of the cognitive process *actually* employed in producing it, it is also a function of processes that could and should be employed.

With these points in mind, we may tentatively propose the following revision of our theory, where we again focus on a base-clause principle but omit certain details in the interest of clarity.

- (10) If S's belief in p at t results from a reliable cognitive process, and there is no reliable or conditionally reliable process available to S which, had it been used by S in addition to the process actually used, would have resulted in S's not believing p at t , then S's belief in p at t is justified.

There are several problems with this proposal. First, there is a technical problem. One cannot use an additional belief-forming (or doxastic-state-forming) process *as well as* the original process if the additional one would result in a different doxastic state. One wouldn't be using the original process at all. So we need a slightly different formulation of the relevant counterfactual. Since the basic idea is reasonably clear, however, I won't try to improve on the formulation here. A second problem concerns the notion of "available" belief-forming (or doxastic-state-forming) processes. What is it for a process to be "available" to a cognizer? Were scientific procedures "available" to people who lived in pre-scientific ages? Furthermore, it seems implausible to say that all "available" processes ought to be used, at least if we include such processes as gathering *new* evidence. Surely a belief can sometimes be justified even if additional evidence-gathering would yield a different doxastic attitude. What I think we should have in mind here are such additional processes as calling previously acquired evidence to mind, assessing the implications of that evidence, etc. This is admittedly somewhat vague, but here again our ordinary notion of justifiedness is vague, so it is appropriate for our analyses to display the same sort of vagueness.

This completes the sketch of my account of justified belief. Before concluding, however, it is

essential to point out that there is an important use of "justified" which is not captured by this account but can be captured by a closely related one.

There is a use of "justified" in which it is not implied or presupposed that there is a *belief* that is justified. For example, if S is trying to decide whether to believe p and asks our advice, we may tell him that he is "justified" in believing it. We do not thereby imply that he *has* a justified *belief*, since we know he is still suspending judgment. What we mean, roughly, is that *he would or could* be justified if he were to believe p . The justificational status we ascribe here cannot be a function of the causes of S's believing p , for there is no belief by S in p . Thus, the account of justifiedness we have given thus far cannot explicate *this* use of "justified." (It doesn't follow that this use of "justified" has no connection with causal ancestries. Its proper use may depend on the causal ancestry of the cognizer's cognitive state, though not on the causal ancestry of his believing p .)

Let us distinguish two uses of "justified": an *ex post* use and an *ex ante* use. The *ex post* use occurs when there exists a belief, and we say of *that belief* that it is (or isn't) justified. The *ex ante* use occurs when no such belief exists, or when we wish to ignore the question of whether such a belief exists. Here we say of the *person*, independent of his doxastic state vis-à-vis p , that p is (or isn't) suitable for him to believe.¹⁷

Since we have given an account of *ex post* justifiedness, it will suffice if we can analyze *ex ante* justifiedness in terms of it. Such an analysis, I believe, is ready at hand. S is *ex ante* justified in believing p at t just in case his total cognitive state at t is such that from that state he could come to believe p in such a way that this belief would be *ex post* justified. More precisely, he is *ex ante* justified in believing p at t just in case a reliable belief-forming operation is available to him such that the application of that operation to his total cognitive state at t would result, more or less immediately, in his believing p and this belief would be *ex post* justified. Stated formally, we have the following:

- (11) Person S is *ex ante* justified in believing p at t if and only if there is a reliable belief-forming operation available to S which is such that if S applied that operation to this total cognitive state at t , S would believe p at t -plus-delta (for a suitably

small delta) and that belief would be *ex post* justified.

For the analysans of (11) to be satisfied, the total cognitive state at *t* must have a suitable causal ancestry. Hence, (11) is implicitly an Historical account of *ex ante* justifiedness.

As indicated, the bulk of this essay was addressed to *ex post* justifiedness. This is the appropriate analysandum if one is interested in the connection between justifiedness and know-

ledge, since what is crucial to whether a person *knows* a proposition is whether he has an actual *belief* in the proposition that is justified. However, since many epistemologists are interested in *ex ante* justifiedness, it is proper for a general theory of justification to try to provide an account of that concept as well. Our theory does this quite naturally, for the account of *ex ante* justifiedness falls out directly from our account of *ex post* justifiedness.

Notes

- 1 "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1967), pp. 335–72; "Innate Knowledge," in S. P. Stich (ed.), *Innate Ideas* (Berkeley: University of California Press, 1975); and "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976), pp. 771–99.
- 2 Notice that the choice of a recursive format does not prejudice the case for or against any particular theory. A recursive format is perfectly general. Specifically, an explicit set of necessary and sufficient conditions is just a special case of a recursive format, i.e., one in which there is no recursive clause.
- 3 Many of the attempts I shall consider are suggested by material in William P. Alston, "Varieties of Privileged Access," *American Philosophical Quarterly* 8 (1971), pp. 223–41.
- 4 Such a definition (though without the modal term) is given, for example, by W. V. Quine and J. S. Ullian in *The Web of Belief* (New York: Random House, 1970), p. 21. Statements are said to be self-evident just in case "to understand them is to believe them."
- 5 Page 22.
- 6 I assume, of course, that "nomologically necessary" is *de re* with respect to "S" and "t" in this construction. I shall not focus on problems that may arise in this regard, since my primary concerns are with different issues.
- 7 This assumption violates the thesis that Davidson calls "The Anomalism of the Mental." Cf. "Mental Events" in L. Foster and J. W. Swanson (eds), *Experience and Theory* (Amherst: University of Massachusetts Press, 1970). But it is unclear that this thesis is a necessary truth. Thus, it seems fair to assume its falsity in order to produce a counterexample. The example neither entails nor precludes the mental-physical identity theory.
- 8 Keith Lehrer's example of the gypsy lawyer is intended to show the inappropriateness of a causal requirement (see *Knowledge* (Oxford: Clarendon Press, 1974), pp. 124–5.) But I find this example

- unconvincing. To the extent that I clearly imagine that the lawyer fixes his beliefs solely as a result of the cards, it seems intuitively wrong to say that he *knows* – or has a *justified belief* – that his client is innocent.
- 9 This definition is not exactly what we need for the purposes at hand. As Ernest Sosa points out, introspection will turn out to be a belief-dependent process, since sometimes the input into the process will be a belief (when the introspected content is a belief). Intuitively, however, introspection is not the sort of process which may be merely conditionally reliable. I do not know how to refine the definition so as to avoid this difficulty, but it is a small and isolated point.
- 10 It may be objected that principles (6_A) and (6_B) are jointly open to analogues of the lottery paradox. A series of processes composed of reliable but less-than-perfectly-reliable processes may be extremely unreliable. Yet applications of (6_A) and (6_B) would confer justifiedness on a belief that is caused by such a series. In reply to this objection, we might simply indicate that the theory is intended to capture our ordinary notion of justifiedness, and this ordinary notion has been formed without recognition of this kind of problem. The theory is not wrong *as* a theory of the ordinary (naive) conception of justifiedness. On the other hand, if we want a theory to do more than capture the ordinary conception of justifiedness, it might be possible to strengthen the principles to avoid lottery-paradox analogues.
- 11 I am indebted to Mark Pastin for this point.
- 12 Cf. *Theory of Knowledge*, 2nd edn, pp. 17, 114–16.
- 13 Cf. "A Causal Theory of Knowing." The reliability aspect of my theory also has its precursors in earlier papers of mine on knowing: "Innate Knowledge" and "Discrimination and Perceptual Knowledge."
- 14 The view that introspection is retrospection was taken by Ryle, and before him (as Charles Hartshorne points out to me) by Hobbes, Whitehead, and possibly Husserl.

- 15 Of course, if people in world *W* learn *inductively* that wishful thinking is reliable, and regularly base their beliefs on this inductive inference, it is quite unproblematic and straightforward that their beliefs are justified. The only interesting case is where their beliefs are formed *purely* by wishful thinking, without using inductive inference. The suggestion contemplated in this paragraph of the text is that, in the world imagined, even pure wishful thinking would confer justifiedness.
- 16 I am indebted here to Mark Kaplan.
- 17 The distinction between *ex post* and *ex ante* justifiedness is similar to Roderick Firth's distinction between *doxastic* and *propositional* warrant. See his "Are Epistemic Concepts Reducible to Ethical Concepts?" in Alvin I. Goldman and Jaegwon Kim (eds), *Values and Morals, Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt* (Dordrecht: D. Reidel, 1978).

How to Think about Reliability

William P. Alston

I

Theories of epistemic justification and of knowledge in terms of *reliability* have been prominent lately. The way it typically goes is that justification of belief is explained in terms of reliability and then having a justified belief that *p* is taken as necessary for knowledge that *p*. Among those who propound theories of this sort are Alvin Goldman,¹ Marshall Swain,² and Frederick Schmitt.³ The basic intuitive idea of a reliabilist theory of justification is that justified belief is belief that results from a reliable belief-forming process.⁴

If *S*'s believing *p* at *t* results from a reliable cognitive belief-forming process... then *S*'s belief in *p* at *t* is justified.⁵

In this paper I will not be concerned with setting out, defending, or attacking a reliabilist account of justification or knowledge. Indeed, I will not be directly concerned with epistemological issues at all. What I will treat is the *concept* of *reliable belief formation*. I am prompted to do this by the fact that among the many criticisms that reliabilist epistemology has evoked, a prominent place is occupied by conceptual criticisms. It has been claimed more than once that the notion of *a belief's being formed by a reliable process* is too indeterminate to permit any satisfactory and objective answer to the question of whether it applies in a given case. If that is so, the theory can't get off the ground. If our grasp of *reliability* is such that we

are, in principle, unable to tell whether a given belief is formed reliably or not, we cannot take the first step toward considering whether, for example, a belief is justified *iff* it is reliably formed. Reliabilist epistemologists, including Goldman and Schmitt, have taken such criticism very seriously and have attempted to respond to them. I find these attempts not to be very successful.⁶ In this essay I hope to do better. My aim is to articulate a conception of *reliable belief formation* that can be used to set out a reliabilist theory of epistemic justification that is at least not internally incoherent.

II

The criticism on which I will concentrate has to do with what Richard Feldman has called the *Problem of Generality*.⁷ The problem arises as follows. A particular belief is generated by a particular psychological process. But a particular process, with a certain spatiotemporal location, is not the sort of thing that can be more or less reliable. Reliability or the reverse attaches only to what is repeatable, to what has, actually or potentially, a number of instances. At least this is true if we are thinking of events or processes, as reliabilists typically are. It is more common in ordinary speech to attribute reliability and unreliability to mechanisms, like thermometers, carburetors, and clocks, or to medicines, or to sources of information, whether persons or books or newspapers. These are all *substances, things* of one kind or another, and here *repeatability* and *instances* are not in the picture. Nevertheless, an analogous point can be made.

Originally published in *Philosophical Topics* 23 (1995), pp. 1-29; reprinted by kind permission of the author.

Though a clock or an encyclopedia or a medicine is not “repeatable,” it is something that can be operated, consulted, or used many times, and so there is something like repeatability here, viz., repeated employments. So the general point is that to be assessable as reliable or the reverse, something must, actually or potentially, provide a range of cases of the appropriate sort. For reliability is always a matter of the incidence of favorable outcomes in a multitude of instances or employments of the item in question. And if there is not at least the possibility of such a multitude, the question of the proportion of favorable cases cannot arise. What counts as a *favorable* case differs, of course, from one application to another. With medicines, it is a certain kind of medical improvement. With clocks, it is registering the correct time. With maps, it is accurate representation of location and distance. For the case at hand – belief-forming processes – it is truth. A reliable belief-forming process is one that usually yields *true* beliefs. And, to get back to the original point, a particular process that takes place at a particular precise time is not the sort of thing that does or does not enjoy a favorable ratio of true beliefs among its products. It occurs just once; the one belief it produces is either true or false, and there’s an end to it. Hence, as is regularly said by both friend and foe, it is a *type* of cognitive process, rather than a particular process (a *token*), that can be assessed for reliability. Consider a particular case of *inferring a generalization from a number of instances*. If we want to ask how reliable that process is, we have to consider the *type* of inference to which it belongs and seek to determine the proportion of true generalizations among the (actual or possible) products of *tokens of that type*.

Before continuing, I had best make it explicit that reliability is a degree concept. An instrument, book, or person can be more or less reliable. Thus when one uses “reliable” and “unreliable” as absolute, yes-or-no terms, one must, if one is talking sense, have some minimum degree of reliability in mind as a necessary condition for speaking of the item as reliable *tout court*. We don’t want to be too rigid about this. One’s conception of reliability, like many other concepts, might be less than fully precise without lacking content. Being reliable might be thought of as “usually” or “most often” exhibiting favorable cases, or as there being a “high proportion” of favorable cases, or as “most” of the cases being favorable, rather than being thought of in terms of a precise minimum

percentage.⁸ I mention this point only to set it aside. Though it is an important point for the development of a reliabilist epistemology, it does not bear on the issues I will be discussing in this essay.

These are the rudiments of the situation. Thus far the difficulty has not shown itself. But it is not far below the surface. If we seek to determine whether a particular belief is justified by determining whether it was produced by a reliable process, the above remarks indicate that we must determine whether the *type* of belief-forming process of which this belief formation is an instance usually does or would generate true beliefs. And the difficulty is that there is no unique type that meets that description. Any particular process is an instance of indefinitely many process types. I will let Feldman make the point.

[T]he specific process token that leads to any belief will always be an instance of many process types. For example, the process token leading to my current belief that it is sunny today is an instance of all the following types: the perceptual process, the visual process, processes that occur on Wednesday, processes that lead to true beliefs, etc. Note that these process types are not equally reliable. Obviously, then, one of these types must be the one whose reliability is relevant to the assessment of my belief. Intuitively, it seems clear that the general reliability of processes that occur on Wednesday or processes that lead to true beliefs is not relevant to the assessment of my belief. The reliability of the visual process or of the perceptual process may well be important.

Let us say, then, for each belief-forming process token there is some “relevant” type such that it is the reliability of that type which determines the justifiability of the belief produced by that token. Thus, the reliability theory can be formulated as follows:

(RT) S’s belief that *p* is justified if and only if the process leading to S’s belief that *p* is a process token whose relevant process type is reliable.

In order to evaluate (RT), we need some account of what the relevant types of belief-forming processes are. Without such an account, we simply have no idea what consequences the proposal has since we have no

idea which process types are relevant to the evaluation of any particular beliefs.⁹

Here is another formulation of the same complaint by Alvin Plantinga.

The main problem, as I see it, still remains. Note first that any particular token – any relevant sequence of concrete events – will be a token of many different types. Consider a specific visual process in Paul, where the input consists in retinal stimulation, let's say, and the output consists, for some particular scene *s* on his television, in his believing that he sees *s*. The process in question will presumably involve a large number of events; it will no doubt include an event consisting in Paul's being appeared to in a characteristic way. Now this sequence of events will be a token of many different types – the cognitive process, the visual process, the cognitive process occurring on a Thursday, the visual process occurring in a middle aged man, the visual process occurring in a middle aged man under such and such lighting conditions, the visual process occurring in a middle aged man when his retinas are being stimulated by light of such and such a character; and many more.

It is these types that are to be evaluated for reliability (since, as we recall, the degree of justification enjoyed by the belief in question is a function of the reliability of the process (type) causing it); but obviously the types may differ wildly among themselves with respect to reliability. Which is the relevant type? Which type is the one such that its reliability determines the justification Paul has for the belief in question?¹⁰

Thus there is no unique type to which a particular process token belongs. Each token belongs to indefinitely many types. Hence to construct a reliabilist theory of justification, we must find some way to pick from this embarrassment of riches a type on the basis of which a judgment of the reliability of this particular belief formation can be made. And, they continue, there would seem to be no satisfactory way of doing this. Here is Feldman's picture of the situation.

In coming up with an account of relevant types, defenders of the reliability theory must be

guided by the following point. If relevant types are characterized very narrowly then the relevant type for some or all process tokens will have only one instance (namely, that token itself). If that token leads to a true belief, then its relevant type is completely reliable, and according to (RT), the belief it produces is justified. If that token leads to a false belief, then its relevant type is completely unreliable, and, according to (RT), the belief it produces is unjustified. This is plainly unacceptable, and in the extreme case, where every relevant type has only one instance, (RT) has the absurd consequence that all true beliefs are justified and all false beliefs are unjustified. We can say that characterizing relevant types too narrowly leads to the "The Single Case Problem."

A very broad account of relevant types of belief-forming processes leads to what we may call "The No-Distinction Problem." This arises when beliefs of obviously different epistemic status are produced by tokens that are of the same (broad) relevant type. For example, if the relevant type for every case of inferring were the type "inferring," then (RT) would have the unacceptable consequence that the conclusions of all inferences are equally well justified (or unjustified) because they are believed as a result of processes of the same relevant type.

The problem for defenders of the reliability theory, then, is to provide an account of relevant types that is broad enough to avoid The Single Case Problem but not so broad as to encounter The No-Distinction Problem. Let us call the problem of finding such an account "The Problem of Generality."¹¹

Feldman then goes on to argue at some length that attempts to solve this problem have not been successful, and he suggests that the prospects for doing better are poor.¹²

III

Now if I were to accept Feldman's challenge on his terms and attempt to respond, I would be engaged in what I said I would not do in this essay, viz., enter into the merits and demerits of reliabilist theories of justification. For to find a way of specifying "relevant types" that avoids both his Scylla and his Charybdis is to find a way of assigning

tokens to types that makes the former line up with intuitive judgments of justification in the right way, and I have disavowed any intention of doing that here. What I will do instead is to challenge the basic assumption that lies behind the challenge, viz., that there are no objective, psychological facts of the matter that pick out a unique type as the one of which a particular process is a token. It is only if that assumption is true that the reliabilist epistemologist is faced with the daunting task of providing a principled way of selecting a single winner from this plethora of candidates. And note that even if she should succeed in that task, she would still be subject to the complaint that the assignment of each token to a unique type has been rigged to fit an antecedent decision as to the epistemic status of the belief, thereby giving rise to the suspicion that reliability is not the most basic determinant of justification after all. Hence a great deal is riding on the supposition that there are no objective, *nonepistemic* facts that determine a unique type assignment for each token belief formation. It is that supposition I will seek to discredit.

Before turning to that, however, I can use purely conceptual, nonepistemic, considerations to show how to avoid Feldman's Scylla, the "Single Case Problem." That can be disposed of just by making it explicit that reliability is *not* a matter of actual track record but rather is a "propensity" or "disposition" notion. To say that a thermometer, medicine, or atlas is *reliable* is not to make a report of the relative frequency of favorable outcomes in the cases in which it has been used up to this point. It may never have been used at all yet, but that doesn't keep it from being reliable or unreliable. We may not be able to tell how reliable it is if it hasn't yet been used, but here as elsewhere it is a great mistake to conflate *X's being P* with our ascertaining, or being able to ascertain, that it is *P*. A thermometer may be perfectly reliable, as we may discover in time, even though it has just rolled off the assembly line and has not yet been put to the test. Nor is the reliability of a device a function of the proportion of favorable outcomes of uses over its entire life history – past, present, and future. An atlas might be perfectly reliable even though no one ever opens it or consults it for any purpose. In this respect reliability functions like other dispositional properties. A rubber band can be elastic even though it is never stretched and never has the chance to manifest that disposition. The applicability of a dispositional term depends

on whether the appropriate manifestations *would* result from the satisfaction of the relevant antecedent conditions in a suitable range of cases, whether or not such a spread of cases, or any cases at all, are ever forthcoming. An elastic substance is one that *would* resume its shape *if* deformed. A responsible person is one who *would* fulfill whatever commitments he or she made, or most of them. A reliable instrument is one that *would* usually deliver favorable results over an appropriate range of cases *if and when* they occur. The same point holds, with appropriate adjustments, for process types. A type of belief-forming process is a reliable type provided that tokens of that type *would* generate true beliefs in a large proportion of a suitable run of cases.

In denying that reliability is a matter of actual frequencies, I am not denying that the most direct way of assessing a device or a process type for reliability is to ascertain the frequency of favorable outcomes in a suitable range of cases. That is not the only way, but it is certainly the most direct way. And if we couldn't use that approach for many dispositional properties, we would not be in a position to employ less direct procedures that require already knowing how to apply other dispositional concepts. Thus, having ascertained by actual-frequency counts that certain human perceptual belief-forming processes are reliable, we thereby have some basis for inferring that processes sufficiently similar to those are likely to be reliable also. But the latter inference requires that we have already done some frequency count spade-work for some processes.

Note that I have said that we need to *make explicit* the point that reliability is a dispositional or propensity notion, rather than a track-record notion. I did not say that we need to *develop* the notion in this direction or *modify* it so that it takes this shape. It seems perfectly clear to me that as we ordinarily use the term "reliable" it functions in the way I have just described. Feldman presents the propensity construal as something a reliability theorist might be driven to in order to answer objections.¹³ And even Goldman is much too permissive on this point.

I have characterized justification-conferring processes as ones that have a "tendency" to produce beliefs that are true rather than false. The term "tendency" could refer either to *actual* long-run frequency, or to a "propensity", i.e., outcomes that would occur in merely possible realizations of the process. Which of these

is intended? Unfortunately, I think our ordinary conception of justifiedness is vague on this dimension.¹⁴

But I don't see how anyone could possibly think that "tendency" might mean "actual long-run frequency." Nor do I see how anyone could reasonably suppose that "reliable" (that's the term we should be discussing rather than "justified") is vague on the distinction between propensity and actual frequency. That's certainly not my sense of the way these terms work.

Before continuing the main line of argument, I had better make explicit an important distinction between two types of belief-forming processes. In "What Is Justified Belief?" Goldman distinguishes between "belief-independent" and "belief-dependent" processes. The former take no doxastic inputs; the latter take at least some. Goldman characterizes reliability differently for the two types, and I will follow him in this. The formulation just given applies straightforwardly to belief-independent processes. In their case, reliability is a simple matter of the proportion of true outputs that would result from an appropriate range of cases. But the latter division is more complicated. There we have to give separate attention to the transfer of epistemic status across the inference and to the status that is being transferred. Goldman handles this by treating separately the reliability of the process and the reliable formation of the output belief.

A [belief-dependent] process is *conditionally reliable* when a sufficient proportion of its output-beliefs are true *given that its input-beliefs are true*.

If S's belief in p at t results ("immediately") from a belief-dependent process that is (at least) conditionally reliable, and if the beliefs (if any) on which this process operates in producing S's belief in p at t are themselves justified, then S's belief in p at t is justified.¹⁵

Dividing the situation up in this way enables us to assess the process for reliability without worrying about the epistemic status of the inputs and then to use that result, along with the epistemic status of the inputs, to give conditions for the belief to satisfy reliabilist conditions for justification. With belief-independent processes there is no need for

this two-stage treatment, since the inputs (experiences) are not the sort of things that have epistemic status.

The application of my point that reliability is a propensity notion to Feldman's worries about "the single case problem" is quite straightforward. Since reliability doesn't hang on actual frequency, there is no excuse for supposing that if a belief-forming process type is exemplified only once and produces a true belief it is perfectly reliable, whereas if it produces a false belief on its only instantiation it is completely unreliable. Its place on the reliability dimension depends on what the proportion of true beliefs *would be* in a suitable range of instantiations, not on what actually results when it is realized.

In contrasting a track-record construal with a propensity construal, I have presented the latter in terms of the frequency that would be displayed by a "suitable" or "appropriate" range of cases. What does it take for a range of cases to be "suitable"? First, they must be sufficiently numerous. What counts as sufficient will vary with different subject matters. We learn from experience how much homogeneity or heterogeneity to expect in a number of cases of a given sort. The more variation we have learned to expect across instances, the larger the sample we need. Since people vary much more than chemical substances, we need a larger sample for opinion or attitude research than for determining atomic weight. Second, the cases must be sufficiently varied along relevant dimensions to rule out, so far as possible, the hypothesis that the results are due to factors other than the characteristic being tested for. If we are testing for arithmetic ability, we wouldn't want to confine ourselves to very easy or very difficult problems; we would want to include a spread of difficulty in the test. Similarly, if we are testing a perceptual belief-forming process for reliability, we would want to vary the cases with respect to such factors as distracting "noise" from other psychological inputs and the degree of discrimination required between inputs close to each other in time and/or space. These two points are applications to this topic of criteria that are applied to any statistical or sampling procedure.

But there is another factor that must be taken into account here. What the last paragraph suggests is that a belief-forming process will be deemed reliable provided its exercise in a suitable number and range of cases would result in a high proportion of true beliefs. But this would seem to

be false for many of the processes we would ordinarily regard as reliable, so long as no restrictions are put on the circumstances in which the process occurs. Many human belief-forming processes would exhibit sharply different degrees of reliability over the whole range of conceivable situations. This may not be true of all. Some have supposed that introspection must yield only truths, no matter what, and some have taken the apprehension of propositions as self-evident to enjoy a like immunity to error. Moreover, if we restrict ourselves, as I have been doing, to assessing inferential processes in terms of *conditional* reliability, then if we ignore Cartesian doubts about the a priori, it seems that in no possible situation would deductively valid inferential processes fail to be conditionally reliable. But look at perception and nondeductive reasoning. For any perceptual mechanism that produces mostly truths in the situations in which it is actually exercised, there are possible situations in which that reliability would be sharply reduced. And not just logically possible situations either. Even if the Cartesian demon is only logically possible, it is well within our powers to arrange environments in which a normal person, utilizing normal mechanisms of perceptual belief formation that serve us well in run-of-the-mill situations, would, usually or always, be led astray. We need only manufacture realistic enough look-alikes, or do something more ambitious with holographs, or something still more ambitious with direct brain stimulation. If the range of cases in terms of which the process is assessed for reliability were restricted to situations like these, it would score very low, even if the score would be high when tested in more familiar situations. To illustrate the problem with respect to nondeductive reasoning, we must have recourse to more *recherché* possibilities. It does seem at least logically possible that the world should be such that the modes of nondeductive reasoning in which we are most confident would fail to produce mostly true beliefs from true premises and so would score low on conditional reliability. Consider induction by simple enumeration. Surely there are possible worlds in which when there is a high proportion of *F*'s in *G*'s in what we take to be a properly constituted sample, it is usually the case that most *G*'s overall are not *F*. If all else fails, we can introduce a Cartesian demon that delights in arranging things so that such inferences are usually, or invariably, frustrated.

Thus we are faced with a question as to the range of situations in which a process must yield mostly true beliefs in order that it qualify as reliable. If we require reliability over all possible situations, we will, at best, be left with a sharply reduced set of reliable human belief-forming habits. And if not that, how shall we demarcate those situations over which the test sample must range in order to give a relevant result?

I can't think of any better answer to this question than the following. The requirement for reliability is that the process would yield a high proportion of truths over a wide range of situations *of the sort we typically encounter*. Obviously, this is far from precise. It doesn't draw a sharp boundary between typical and atypical. Moreover, it leaves open the possibility that the boundary, such as it is, can shift over time. What was atypical up to now may become typical with cultural, technological, or other changes. However, I believe that this suggestion has the right kind and degree of sloppiness for the concept of reliability we want for epistemic purposes. It does unequivocally rule out clearly atypical situations – Cartesian demons, brains in vats, and the like.¹⁶ And it makes a judgment of reliability dependent on our actual situation as human beings in the environments in which we actually find ourselves. This is what we need to capture the intuitive notion of reliability that is involved in reliabilist epistemologies. If I claim that my thermometer is reliable, it is no refutation to point out that it would not give an accurate reading on the sun. Similarly, if I claim to be able to accurately determine, by vision, when I am standing in front of a beech tree, it would be no refutation of that claim to point out that I could not do this if I were receiving direct stimulation of the visual cortex in a physiological laboratory. When I make a judgment of reliability – whether for an instrument, a documentary source, a psychological mechanism, or whatever – I have in mind, at least implicitly, a range of situations with respect to which the claim is being made. What happens outside that range is simply irrelevant to the claim.

IV

Now I can return to my central task – challenging the assumption that there are no nonepistemic facts that determine a unique type to which a

given belief-forming process belongs. Of course, in a way I agree that a given process token does not belong to only one type. With a process token, as with any other particular, any of its properties can be said to be correlated with a type to which it belongs; and its properties are indefinitely numerous. Nevertheless, some types, in this maximally generous logical sense, are ontologically rooted, fundamental, or important in ways indefinitely many others are not. Even if it is true that you and I belong to indefinitely many classes, such as *objects weighing more than ten pounds*, *objects that exist in the twentieth century*, *objects mentioned in this paper*, etc., etc., it is still the case that membership in the class of human beings is fundamental for *what we are* in a way those others are not, just because it is the *natural kind* to which we belong. I shall suggest that something analogous is true of belief-forming processes – that there are fundamental considerations that mark out, for each such process token, a type that is something like its “natural kind,” and hence that in thinking of belief-forming process types, we are not awash in a sea of indeterminacy as Feldman and company suppose.

To carry this through we will have to think of belief-forming processes in a somewhat different way than is common with both friends and foes of reliabilism. First, we will have to draw boundaries around the process more narrowly than is often done. Most discussions of the current issue are not very specific about the extent of the processes under discussion. One crucial question concerns whether we should think of the process as including everything in the causal ancestry of the belief or only some segment thereof. In particular, should we include events outside the organism or limit them to certain intra-organismic, or intrapsychic, events that function as proximate causes of the belief? Critics typically follow the lead of some reliabilists in leaving this hazy. As an example of the latter, Armstrong’s version of reliabilism runs something like this. *S*’s belief that *p* is reliably formed provided that *S* has some property, *H*, such that it is nomologically necessary that if a subject that is *H* forms a belief that *p*, that belief is true.¹⁷ This formulation is not explicitly in terms of processes, but it could be rewritten in those terms, as follows. *S*’s belief that *p* is reliably formed provided the process that generated that belief has some property, *H*, such that it is nomologically necessary that any belief that is generated by a process with that property is true. This

obviously puts no restriction on the extent of the process. Any characterization of the process that will yield the nomological necessity in question will ensure reliability of belief formation.¹⁸ As for the critics, we find Feldman and Pollock considering perceptual processes that include factors outside the subject, such as distance from the object and lighting conditions.

However, our paradigm reliabilist, Goldman, is more specific.

In addition to the problem of “generality” . . . there is the problem of the “extent” of belief-forming processes. Clearly, the causal ancestry of beliefs often includes events outside the organism. Are such events to be included among the “inputs” of belief-forming processes? Or should we restrict the extent of belief-forming processes to “*cognitive*” events, i.e., events within the organism’s nervous system? I shall choose the latter course, though with some hesitation. My general grounds for this decision are roughly as follows. Justifiedness seems to be a function of how a cognizer deals with his environmental input, i.e., with the goodness or badness of the operations that register and transform the stimulation that reaches him. . . . A justified belief is, roughly speaking, one that results from cognitive operations that are, generally speaking, good or successful. But “*cognitive*” operations are most plausibly construed as operations of the cognitive faculties, i.e., “information-processing” equipment *internal* to the organism.¹⁹

This seems to me just the right thing for a reliabilist to say on this point. If the epistemic status of a belief is a function of the reliability of the process that generates the belief, it is the reliability of the *psychological* process that is crucial. Looking at perceptual belief formation, no matter how exemplary the path of the light rays from the surface of the perceived object to the retina, and no matter how finely tuned the neural transformations involved in the pathway from the eye to the brain, if the belief is not formed on the basis of the conscious presentation (and/or its neural correlate) in a truth-conducive way, the belief will lack the epistemic desideratum that is stressed by reliabilism. It is that final step that is crucial. I note that by the time he came to write *Epistemology and Cognition* Goldman had lost the “hesitation” of which he speaks in the above passage. There he

deals with *psychological* processes from the outset. See particularly chapter five.

But this limitation of belief-forming processes to the psyche does not significantly reduce the embarrassment of riches where types are concerned. A purely cognitive belief-forming process will also be of indefinitely many types, including such undesirable intruders as *happening on a Wednesday* and *generating a true belief*. Moreover it will belong to types of all levels of generality. If it is a visual belief-forming process, it will be of the type *forming a belief on the basis of such and such a kind of visual presentation, forming a belief about a tree, forming a belief about something in the vicinity, forming a belief on the basis of vision, forming a belief on the basis of perception*, and so on. We are still drowning in indeterminacy.

But decisive help is near. In fact, the germ of it is to be found in the very essay of Goldman's from which I have been quoting, though, as we shall see, he fails to take advantage of that idea to solve the generality problem.

We need to say more about the notion of a belief-forming "*process*." Let us mean by a "*process*" a *functional operation* or procedure, i.e., something that generates a *mapping* from certain states – "inputs" – into other states – "outputs." The outputs in the present case are states of believing this or that proposition at a given moment.²⁰

The crucial point here is that every belief formation involves the activation of a certain psychologically realized *function*. That activation yields a belief with a propositional content that is a certain function (the function the psychological realization of which is activated here) of the proximate input. The function involved will determine both what features of the input have a bearing on the belief output and what bearing they have, i.e., how the content of the belief is determined by those features.²¹ In order to bring our talk of reliability of belief closer to such paradigm subjects of reliability attribution as thermometers, medicines, and atlases, let's say that a psychologically realized belief-formation function constitutes a psychological *mechanism*. If you don't like this terminology, either because it sounds too "mechanistic" or because it threatens to populate the mind-brain with unmanageably many separate black boxes, we can just as well use other terms. We can think of the psychological realization of the function as a

habit of forming a belief with a certain propositional content that is a certain function of certain features of the input or as a *disposition* to do that. Or if you prefer *act-psychology*, we can think of the subject's having the power to "take account" of certain features of inputs and, on the basis of that *taking-account*, form a belief with a content that is a certain function of those features. The common thread running through all this is that it is part of the constitution of the psyche to be so disposed that upon being presented with certain kinds of input a belief is generated with a content that is a certain function of certain features of that input. Remembering the variety of terminology available, I will in the ensuing oscillate between speaking of *habits* and of *mechanisms*.

Let's descend from these high levels of generality and look at some examples. Consider the formation of a visual perceptual belief that a maple tree is in front of one. The input will be a visual "presentation" of a certain sort, one that involves the perceived object's *looking* a certain way.²² The mechanism that is activated will take account of certain phenomenal features of the presentation, while others will play no role. Certain shape features, certain color features, the spatial distribution of variously colored regions, and contrasts with the surrounding field will be "picked up" by the mechanism, while others will be ignored. As for the latter, many details of the presentation could have been different without changing the content of the belief generated. The tree could have looked larger or smaller, the bark could have looked rougher or smoother, and so on. It all depends on what function is operative. Obviously, if the function were one that delivered a belief about size or finely discriminated bark texture, the features that are irrelevant to whether one believes merely that there is a maple tree in front of one would have been relevant.

In this example I assumed that only the visual experience (presentation) functioned as an input. It is widely held that in every case of perceptual belief formation other beliefs of the subject play a role in shaping the doxastic output. Whether or not that is so, it is clear that in many cases the input is partly doxastic. As an example take an "individual recognition" case, rather than a "kind recognition" case like the previous one. Upon seeing a house I form the belief that it is your house. Actually there are a number of houses in the world that look just like yours, on a fleeting glance. But yours is the only one on this block of

this street in this town that looks like yours. Hence, let us say, my identification of the house as yours (my formation of the belief that this is your house) is influenced by my knowledge (belief) that I am on this block of this street in this town, as well as by features of the visual presentation.

These examples may give the impression that the function involved in any perceptual belief formation is extremely specific, relating very detailed features of experiential input to a unique belief content. In view of the heavy weather made by Feldman and others about the problem of navigating between too specific and too general a relevant process-type, it will be pertinent for us to consider for a moment where psychologically realized belief-forming functions stand on that dimension. In a word, the answer is that the operative function can be of various degrees of generality. Sticking with perceptual belief formation for the moment, the function could be so specific as to take only precisely defined experiential features as input and issue only beliefs with a particular propositional content. That is the way I was thinking of the above examples. But there are other possibilities. Consider attributions of color to perceived objects. Here we have what is plausibly regarded as a single function that maps the position of certain aspects of a visual input on several color-relevant dimensions onto a belief that the object seen is of a certain color. Here we can be confident that the function is unitary because of the systematic character of the mapping. However I take this to be the exception rather than the rule for perceptual belief formation. For most propositional contents we don't have the possibility of systematic mapping that we have for color. Hence I will generally think of perceptual belief-forming functions as maximally specific.

Now let's look at a couple of inferential belief formations. It should be clear that on the approach I am suggesting the "functions" involved in deductive inference will be principles of inference. Think of a case of hypothetical syllogism. I reflect that if I refuse your request for a raise you will quit your job, and if you quit your job I will be unable to find a replacement in time for a big contract that is impending. I, naturally, infer that if I refuse your request for a raise I will be unable to find a replacement in time for that contract. The function that yields a belief that is related in that way to the input beliefs is a psychological realization of the principle of inference called "hypothetical syllogism." Obviously, we shouldn't suppose that

only valid principles of deductive inference are psychologically realized. Some people, unfortunately, are so constituted as to regularly form beliefs related to the input beliefs in the pattern known as affirming the consequent. Such a function would yield the belief that *it rained last night* from the input beliefs that *if it rained last night the grass would be wet and the grass is wet*.

Nondeductive inference presents a more complicated picture. If I arrive at a generalization from knowledge of various instances, I typically take into account not only the instances in question but also facts about the subject matter that indicate what kind of sample I need in order to justifiably move to the generalization. As I pointed out above, some ranges of fact are more homogeneous than others and require less size and variety in the sample for a sound generalization. To be sure, we must keep reminding ourselves that we are dealing here with what functions are actually psychologically realized, not only the ones that are reliable. But I think we may safely assume that in many cases considerations of the sort just mentioned figure in the input to an induction by simple enumeration. Again, consider inference to the best explanation. If I am trying to explain the presence of a pool of water on the floor of my basement, the relevant inputs will include not only beliefs about what the state of affairs is currently in my basement but also a list of possible causes of the water's being there, and considerations that bear on the likelihood of each of these causes having been operative.

The reader will, no doubt, have noticed that in the above I switched at a certain point from speaking of belief-forming *processes* to speaking of belief-forming *habits* (mechanisms, dispositions...). This may give the impression that my *activation-of-realized-functions* construal of belief formation is simply a different approach to the matter that has no relevance to the approach in terms of the reliability of *processes* of belief formation. But that is not the case. Habit and process are two aspects of the same phenomenon – belief formation. I am still speaking of belief-forming processes, but I have, following Goldman, made the conception of such processes more specific by thinking of such process as the operation of a *habit* such as I have been describing. Whenever a belief-forming habit that involves a certain input-belief output function is in operation, there is a *process* of belief formation that consists of the input giving rise to the output in accordance with the function. It is just that instead

of thinking, very unspecifically, of the process as whatever is involved in the causal history of the belief, I have limited the perspective to the proximate stage of that process and construed it in accordance with the input-function-output model.

V

The time has come to apply all this to the generality problem. The application is very simple; it has probably already leaped to the eye of the reader. *The function determines the relevant type.* I form the visual belief that a car is parked in front of my house. What type of belief-forming process is such that *its* reliability is crucial for the epistemic status of that belief, according to reliabilist epistemology? The type that is defined by the operative function, viz., *belief formations that proceed in accordance with the function that is involved here.* In other words, the particular process, by virtue of being a *functional* mapping of input features onto output content, has a *built in* generality that is provided by the function. The function *is* something inherently general, and it defines the type the reliability of which, according to reliabilism, is crucial for the justificatory status of the belief in question.

Let me say a bit more about just why the type defined by the operative function is the one to consider if we are interested, from the standpoint of reliabilist epistemology, in how reliably this particular belief was formed on this occasion. The type determined by the function has this special status just because it reflects or embodies the actual dynamics of the process, what is responsible for *this* belief with *this* content being formed on *this* basis. Hence if we assume, as reliabilist epistemology does, that the epistemic status of a belief is a function of its proximate causal history, of what led to its acquisition, then it is this type the reliability of which should be considered. To be sure, that assumption can be contested, but that issue lies outside the bounds of this paper, which is concerned with clearing up conceptual problems in reliabilist epistemology.

To return to the visual belief that a car is parked in front of my house, it would be an immense labor to spell out the function involved here in complete detail, since it would mean enumerating the features of the visual presentation that led to my identifying what I saw as a car parked in my driveway. (Specifying the output side of the function –

the belief content – is no problem.) I can, of course, identify the experiential input accurately, though not analytically, by using the content of the belief output. I can say that the mechanism generated the belief that there is a car parked in my driveway on the basis of an input that consisted in *my being appeared to* “*car-parked-in-my-drivewayly*,” or in *its being just as if I were seeing a car parked in my driveway*, or in *an object’s looking like a car parked in my driveway*. But whether we identify the perceptual input in analytically illuminating terms or in output-dependent terms, the basic point is the same. The type of process the reliability of which is relevant to the epistemic assessment of the belief is the one defined by the function, which is in turn defined by a certain way of going from input features to output features. The question of reliability that is of significance for the epistemic status of this belief is the question of how reliable *this* habit, the one defined by *this* function, is. The question is as to the proportion of true beliefs in the outputs of *this* habit over a sufficiently large spread of appropriately varied cases, in typical circumstances. In other terms, the question is as to the reliability of forming a belief *like this* on the basis of a perceptual presentation *like this*, where the relevant respects of likeness are determined by the constitution of the function realized in *this* mechanism.

As we have already seen, this point can be made more sharply with inferential belief-forming mechanisms, since there the function can be more easily specified, at least for deductive inferences. If I form the belief that *Jim will come to the party* on the input of the beliefs that *Jim will come if he is well* and that *he is well*, then the belief formed here will be reliably formed provided my principle of inference is conditionally reliable – such as to lead to truths from truths, and the input beliefs have been reliably formed. The first of those two conditions will be realized if the principle that constitutes the function realized in the mechanism that gave rise to this belief is *modus ponens*. In that case the relevant type is such that not only is there a large proportion of true beliefs generated by processes of that type that take true beliefs as inputs; the process invariably yields true beliefs from true belief inputs.

Nondeductive inference, as we have seen, presents a much messier picture. Here, in order to be significantly reliable, a function must be sensitive not only to formal properties of the argument, e.g., that it is a generalization from instances, but also to

a variety of more substantive considerations – the character of the sample if it is a case of generalization, various bits of relevant background knowledge, the field of competing explanations if it is an explanatory inference, and so on. Still, whatever the complexities of nondeductive-inference functions, the basic point is the same. What the epistemic status of the particular belief depends on, according to reliabilism, is the conditional reliability of the operative mechanism, i.e., the extent to which the function realized by that mechanism would yield true belief outputs from true inputs in a suitable spread of cases, plus the epistemic status of the input. Or, to put it in terms of processes, the crucial issue is the conditional reliability of the process of going from input to belief output in accordance with that function, along with the epistemic status of the inputs.

So when we think of the reliability of belief-forming processes in this way, we are no longer faced with an indefinitely large multiplicity of types among which we have to find some way of making a choice. With the illumination shed by this way of construing belief-forming processes, the “Problem of Generality,” as construed by Feldman, Plantinga, and Pollock, dissipates like mist before the morning sun. To be sure, it is still true that a particular process token is an instance of an indefinite variety of process types, including countless silly ones like *processes that take place on Wednesday* and *processes that take place in the shower*. But now that we think of belief-forming processes as the functioning of a mechanism that embodies an input–output function, we can ignore all that. The function defines the epistemologically relevant type, and we can forget about the rest.

VI

But though the plurality of candidates for the position of relevant type has been greatly reduced, aren't there still serious alternatives among which this “functional” conception of belief-forming processes does not choose? Are there not many different ways of carving the psyche up into distinct mechanisms or habits of belief formation? Depending on which of these we pick, we will end up with one or another assignment of a particular process to a general habit, habits that differ as to degree of reliability. Go back to my formation of the visual belief that there is a maple tree in front of me. Can't I think of the habit involved as

one of (a) coming to believe that there is a maple tree in front of me on the basis of a visual presentation with such and such features, or (b) coming to believe that there is a tree in front of me on the basis of sensory experience, or (c) coming to believe that a certain plant is spatially related to me in a certain way, on the basis of sensory experience, or...? That is, it looks as if I may think of the habit activated as possessing any one of widely varying degrees of generality. And it may well be that these habits differ in degree of reliability. Presumably the reliability of the habit of forming maple-tree-in-front-of-me beliefs on the basis of visual presentations with just these features is much higher than the reliability of the more general habit of forming tree-in-front-of-me beliefs on the basis of some sensory experience or other. And does this not mean that we have still failed to pick out a unique relevant type?

NO. At least, we are not still confronted with that problem if the assumptions I have been making are warranted. To properly respond to the above objection I need to distinguish the input and output sides. For the latter we can safely assume that the content can be determinately specified. We only have a case of belief formation if we have a belief, one with a particular propositional content. That gives us our starting point. So if we have a case in hand at all, we are not free to specify the output end of the function in various ways. If the belief the formation of which we are considering is a belief that there is a maple tree in front of me, then that ties down the output side, and there is no scope for choosing between different ways of characterizing it. No doubt, that belief belongs to various wider classes of beliefs – *that some plant is in front of me, that something is in front of me, that some plant is spatially related to me*, and so on. But none of these specifications gives us the precise or full account of the content of the belief in question. Hence whatever mechanism produced this belief can't be one that embodies a function for going from a certain input to a belief with one of those more general contents, for that is not the belief that was actually produced in this instance. So we need not worry about a latitude in how the output side of the operative function is specified.

The input side and the function involved are a bit trickier, because they, especially the function, are not so open to view. Just what features of the input are picked up by the mechanism and just how the function “uses” them to determine features of the output (i.e., just what the function is)

is not so obvious. At this point it becomes clear that the current objection forces me to become explicit about a basic assumption of my approach to the problem, viz., its *psychological realism*. I assume that there is always (almost always?) a single answer to the question, "Just what mechanism, embodying just what function, was operative in the generation of this belief?" I assume that just one way of generalizing from this particular sensory-input-belief-output relationship reflects the actual psychological dynamics of the situation. When I look out the window and form the belief that there is a maple tree there, there are, in the abstract, many functions that would yield a belief with that content from a visual presentation of that sort. The transition might be based on the leaf shape, the overall shape of the tree, the color of the leaves, the character of the bark, the size, etc., or some combination of such features. But I am assuming that in that case only one of these possibilities is realized. The mechanism that was operative embodied one of these ways of taking certain features of the concrete input rather than others as the ones that yield a belief with that content. Again, when my beliefs that John will come to the party if well and that John is well yield the belief that John will come to the party, there are many abstract possibilities as to the principle of inference involved. The mere fact that the inference exhibits a modus ponens form does not guarantee that this is the principle that was psychologically operative. I could have been utilizing a function that yields that belief on the basis of any beliefs about John, or on the basis of any set of beliefs one of which is a conditional, or... Nevertheless, according to my psychological realism, exactly one of those possibilities is realized in this case. And whichever one is realized, it is the reliability of that function (or of the correlated mechanism or process) that is crucial for the epistemic status of the belief.

Like any form of realism this one can be opposed. One can doubt or deny that the psyche really is determinate in this way; and one can hold that we are therefore free to make any of indefinitely many choices in picking a determinate function for purposes of assessing a given belief for degree of reliability. And here as elsewhere epistemological motives for antirealism are prominent. It may be claimed that we lack the access to the details of cognitive processes that would be required to determine in each case just exactly what function is operative. One of

Plantinga's objections to reliabilism is along these lines.

Indeed, if, as Goldman suggests, the relevant type must be specified in psychological or physiological terms, we won't be able to specify any such types at all; our knowledge is much too limited for that.²³

I can't enter into a full-dress defense of psychological realism in this essay. The main point I am concerned to make at this point is that the viability of a reliabilist theory of justification or knowledge hangs on the viability of psychological realism. If there is not an objective fact of the matter as to what input-output function is utilized in a given belief formation, then reliabilists are helpless before the Problem of Generality, and they may as well pack their bags and go home. This is why I said that the issues raised by the current objection are so crucial. They go straight to the heart of what it takes for reliabilism to be a real possibility for epistemology.

But short of a full-dress defense I will say this. First, we should not unduly inflate the epistemological requirements for psychological realism about belief-forming mechanisms. To be sure, if our position vis-à-vis such mechanisms were one of total ignorance, the game would not be worth the candle. But we should also be alive to the point that it is highly reasonable to suppose that there are many objective facts we will never know about, and even facts we are incapable of knowing about. Details of the past history of humanity, the earth, the solar system, and, more generally, the universe present many examples of this. Hence human epistemic incapacities vis-à-vis X 's are not necessarily fatal to realism about X 's.

My second point is that our cognitive access to belief-forming mechanisms is not as scanty as my antirealist opponent makes it out to be. Although we, obviously, can't peer into a psyche with some instrument and observe the little input-output functions doing their thing, we are in a similar situation with respect to many other matters about which we know something; and the approaches we use in those cases are available here too. When it is a question of what function was operative in my own case, I often have a "participant knowledge" of this. Although my knowledge of my own input-output mappings is far from ideally complete, and although I am not immune from error in such matters, I typically

can, by reflecting on what is going on, gain some significant degree of insight into what it was that led me to form a certain belief. In the perceptual cases, even if I can't spell out in detail the "atomic" perceptual cues that enabled me to recognize what I was looking at as a maple tree, still I have some insight into the "look" of the object that tipped me off; and I can recognize relevantly similar looks when they recur. In inferential cases I can often, if I am analytically inclined, formulate the principle(s) of inference on which I was relying; or at least I can recognize relevantly similar inferences when they occur. As for third-person cases, if I want to know what function was operative in Sam's inference about John's coming to the party, I can put him to the test in other cases with and without a *modus ponens* structure, and with or without the other features mentioned above. In that way I can try to find consistent patterns in the way he draws conclusions; and although success is not guaranteed, I might reach some fairly solid results. The perceptual cases are again more difficult, but there too I can vary input-output relationships in such a way as to give myself a chance to find consistent, fairly stable patterns of relating perceptual-presentation features to belief content. These techniques yield less than maximally conclusive results for several reasons. For one thing, a person might be utilizing a given function on one occasion, even though she doesn't do so regularly, often, or consistently. Habits of belief formation, like other habits, can be more or less stable, and they definitely are subject to change. For another thing, no matter how many competing hypotheses we have eliminated as to what the function is on a given occasion, there are always more looming on the horizon. Nevertheless, by using techniques such as these, we are considerably better off than blankly ignorant as to what function is operative in a given case of belief formation.

Another point about psychological realism. I don't want to overstress the determinacy and precision of belief-forming functions or, for that matter, belief contents. Indeterminacies to which all psychological states and operations are heir are to be found here as well. Some, or all, of my belief-forming habits may be such that there is some leeway as to exactly where certain perceptual features must be on relevant dimensions (of size, color, pitch, etc.) in order to generate a certain belief content. Even if the input sensitivity is perfectly precise, there may be some looseness in the

way in which input features determine belief content. Exactly the same features might at one time yield the belief that a maple tree is present and at another time (with the "same mechanism" operative) yield the more specific belief that a sugar maple is present. Indistinguishable samples might on different occasions yield general beliefs with somewhat different statistical parameters or somewhat different degrees of confidence. And so on. The psychological realism I espouse is committed only to a degree of determinacy of belief-forming mechanisms that is sufficient to make it worthwhile invoking them in thinking about the reliability of belief formation as well as many other matters.

Here are two other complexities that would have to be recognized in an adequate theory of belief-forming processes and an adequate reliabilist epistemology. First, more than one habit might be involved in a particular belief generation. My belief that it is my wife's car that I see parked in our driveway might be generated both by a perceptual mechanism that takes account of features of my current visual experience and by an inferential mechanism that takes as input the belief that she told me when I left in the morning that she would be at home all day. I don't think this kind of overdetermination poses any special difficulties for cognitive psychology, but it does require the reliabilist epistemologist to make a decision as to which mechanism is such that its reliability is crucial for the justificatory status of the belief. As I have said, I am not, in this paper, setting out to develop a reliabilist epistemology. But I will say, in passing, that if each process would have been sufficient by itself to produce that belief, it would seem reasonable for the reliabilist to hold that the belief is justified provided either of the processes is sufficiently reliable.

The second complication is this. I have been talking as if every belief is generated by a single momentary input-output mapping. But, as we all know, some beliefs are arrived at only after a more or less extended period of deliberation, search for evidence or reasons, weighing considerations pro and con, and so on. How are we to fit that sort of thing into the picture I have been developing? Here I believe that it is primarily the psychologist, rather than the epistemologist, who has additional work to do. In developing the psychology of belief formation, the cognitive psychologist has to decide how to represent the structure of these extended deliberative processes. For one thing, more than

input-belief output mappings are involved. The searches for relevant evidence and weighing of pro and con considerations are processes of a different character. Perhaps the thing to say is that the belief-forming process occurs only at the end of the deliberation and that when it occurs it is of the simple, momentary sort of which I have been speaking. Or perhaps some other construal would be preferable. In any event, I am happy to leave this issue to the cognitive psychologist. So far as I can see, a reliabilist epistemology could work with whatever account seems best from the standpoint of psychological theory.

Interestingly enough, Goldman, who, as we have seen, proposed the *input-output function* conception of belief-forming processes, has failed to take advantage of this idea to solve the generality problem. Here is what he says about that problem in the very essay in which that functional construal was advanced.

A critical problem concerning our analysis is the degree of generality of the process-types in question. Input-output relations can be specified very broadly or very narrowly, and the degree of generality will partly determine the degree of reliability . . .

It is clear that our ordinary thought about process-types slices them broadly, but I cannot at present give a precise explication of our intuitive principles. One plausible suggestion, though, is that the relevant processes are *content-neutral*. It might be argued, for example, that the process of *inferring p whenever the Pope asserts p* could pose problems for our theory.²⁴

There is no hint that the identification of the function involved in a particular belief acquisition itself serves to define the relevant type of process. The discussion in Goldman's *Epistemology and Cognition* goes beyond the above remarks, but, aside from using the propensity conception of reliability to dissolve the Single-Case Problem, the additional suggestion amounts to the following.

But how is it determined, in each specific case, which process type is critical? . . . Let me advance a conjecture about the selection of process types, without full confidence. The conjecture is: the critical type is the *narrowest* type that is *causally operative* in producing the belief token in question.²⁵

But if my remarks in this essay are on target, there is only one "type," i.e., only one realized function, that is *causally operative* in a given case, apart from cases of overdetermination. The function, and the process type it defines, can, of course, be characterized in various ways, as we have seen, because it has various properties, intrinsic and relational, beyond the constitution of the function that specifies its epistemically relevant character. But when we are interested in knowing what kind of process to check for reliability in order to assess the belief epistemically, the realized function that was actually operative in the belief generation gives us a unique answer to that (if we know enough about the particular case to take advantage of this possibility).

Why haven't philosophers seen this, even philosophers like Goldman who realize that a belief-forming process is, essentially, the operation of a realized input-output function? Again, I suspect that the reluctance to be this psychologically realist plays a major role here. If one simply talks of "processes," each of which belongs to indefinitely many classes or types of processes, one is not making such strong psychological assumptions. But, as I said above, if one is not prepared to be that realist about the psyche, one should quit trying to be a reliabilist in epistemology.

VII

This completes my central task in this paper – exhibiting a psychological, nonepistemic basis for identifying a unique type in terms of which a particular case of belief formation can be assessed for whether the belief was reliably formed. It is not the case that all type assignments based on properties of the process are on a par. In every case there is one that reflects the actual dynamics of the belief formation, and, naturally, it is this one to which we should look if we are interested in how that particular belief formation stacks up with respect to reliability.

I have said more than once that it is no part of my intention here to enter into the issues that arise when one tries to build on this result to develop a reliabilist account of justification and/or knowledge. Nevertheless, the above results can be used to throw light on some further objections to that enterprise. In this section I will say a few words along those lines about what Feldman calls the "No-Distinction Problem."

You will remember that Feldman presents that problem as follows.

A very broad account of relevant types of belief-forming processes leads to what we may call "The No-Distinction Problem." This arises when beliefs of obviously different epistemic status are produced by tokens that are of the same (broad) relevant type. For example, if the relevant type for every case of inferring were the type "inferring," then (RT) would have the unacceptable consequence that the conclusions of all inferences are equally well justified (or unjustified) because they are believed as a result of processes of the same relevant type.²⁶

Of course the reliabilist will seek to make type assignments such that it is not implausible that all beliefs produced by tokens of a single type have the same epistemic status. The question for me is as to whether epistemically relevant types picked out in the way I have been suggesting pass the no-distinction test. Are actual belief-forming functions sufficiently homogeneous epistemically? I am not going to try to answer that question in this paper. Instead I will point out how the reasons given by Feldman and Plantinga for doubting that any types satisfy this requirement can be seen to fail if we restrict ourselves to psychologically realist types. And I will end with a few considerations that suggest that the belief-forming functions in the human psyche might well be epistemically homogeneous. In this discussion I will assume that the functions in question are maximally specific, in that any difference in input that is registered by the function indicates a different function.

One case used by Feldman, following Goldman, has to do with visually recognizing a mountain-goat. Feldman has been making the obvious point that so broad a type as *visual belief formation* fails the no-distinction test. He then cites Goldman's statement that seeing a nearby object is a different process from seeing a distant object. Feldman supposes that this difference has to be drawn in terms of external conditions of observation.²⁷ But that is not so. The relevant features of the sensory experience will obviously be different in the two cases, and so a different function will be activated. At a later stage he considers borderline cases such as often confront an umpire calling balls and strikes. "Some of the objects the person sees may clearly have the property in question while others

do not. As a result, some beliefs to the effect that the object has the property may be better justified than others."²⁸ And so the products of "the process" are epistemically heterogeneous. But, again, it is plausible to suppose that the subject is sensitive to differences between cases that are visually clear and cases that are not. Hence there will be at least two different functions involved.

I find that all of Feldman's cases can be plausibly disposed of in this way, once we think of relevant types in terms of the input-output functions. Of course, this doesn't show that all belief-forming functions are epistemically homogeneous, any more than Feldman has shown that many of them are not.²⁹ As previewed, I will end by presenting some considerations that are relevant to the consideration of this question.

First it is clear that there is no possibility of epistemic heterogeneity for deductively valid inferential functions, like *modus ponens*, that are logically guaranteed to yield a true conclusion from true premises whatever the subject matter. Outside that realm we don't have the same sort of guarantee. But it still seems plausible that the kinds of functions we have been discussing don't suffer from epistemic heterogeneity. As for non-deductive inference, if we were to try to treat this formally on the model of deduction, we would run into the No-Distinction Problem. Think of the formal inductive function – producing the belief that *Most F's are G's* on the input of one hundred *F's* that are *G's* without any *F's* that are not *G's*. Obviously this works better with some *F's* and *G's* than with others. It gives us a higher reliability if *F* is *golden retriever* and *G* is *affectionate* than it does if *F* is *book* and *G* is *has a red cover*. But it seems clear that the nondeductive inference functions generally internalized are not of this sort. They are, rather, designed to be sensitive to peculiarities of the subject matter as well as to the kind of inference involved. Again, an "argument-to-the-best-explanation" function will be sensitive to a good deal of specific information about the explanandum as well as about the competing explanations and the factors that affect their comparative assessment. And as for perceptual belief functions, so long as they restrict inputs to one set of experiential features, perhaps together with relevant background beliefs, and involve a single way of mapping these onto belief output content, there would seem to be no way in which some subset of their uses would involve a significantly higher proportion of true beliefs than others.

But since, as I have already been at pains to point out, our knowledge of the details of human belief-forming functions is imperfect at best, we cannot be sure that there are no functions that founder on the No-Distinction Problem. Perhaps some, or even many, people instantiate purely formal nondeductive inference functions that are defective in this way. Moreover, as the editor of this issue pointed out to me, even if no human belief-forming functions are epistemically heterogeneous, there are surely possible subjects that instantiate such functions. And doesn't a reliabilist epistemology aspire to handle possible as well as actual cases? Leaving that last issue aside, I will conclude by discussing the possibility of a single humanly realized perceptual function that fails the no-distinction test by virtue of delivering a belief with a certain propositional content on any one of a variety of inputs.

Think of the familiar point that there are various ways which the members of a kind look. Consider "dog," "true," "pine tree," or "house." There is an enormous diversity in the kinds of perceptual presentation that a particular person will regularly take as indicating that a perceived object is a dog or a tree or a house. Not all dogs or all houses look just alike, not by a long shot. And it may well be that my identification of something as a dog by the use of one of these component inputs is much more reliable than others. Some doggy looks are more similar to wolf looks or fox looks than other doggy looks. Some pine trees are harder to distinguish from other conifers than other pine trees. And if it is correct in cases like this to count all my ways of going from the look of *X* to the belief that *X* is a dog, or a pine tree, as exercises of a single habit, then if we determine the reliability status of a particular belief by determining the reliability of the habit the activation of which gave rise to that belief, it looks as if we are stuck with a not inconsiderable number of cases in which the epistemic assessment of the particular belief will be indeterminate because of the heterogeneity of the habit that produced that belief.

But this is a serious worry only if it is a serious possibility that there are unitary belief-forming habits of this sort. And that is dubious. It is by no means obvious, from introspection and other ordinary means of access, just what to say about such cases. Where there is a single unified functional relationship that we can formulate and that is plausible to regard as the content of an actual belief-forming habit, as with perceptual color attri-

butions, then we have an obvious reason for attributing a belief-forming habit that embodies that rather broad function to people. But here we have no such reasons available. Why should we suppose that anyone possesses a *single* function that includes all her ways of perceptually identifying a dog or a pine tree. If we have any basis for doing so, it will be a more *recherché* theoretical basis. And theoretical considerations would seem to tend in the other direction. After all, the only reason for supposing there to be a unified function here is the commonality in the content of the belief output. Aside from the fact that the inputs include features of perceptual presentations, that is the only thing that stretches over the whole territory. The particular experiential features vary widely, and the ways in which these features are mapped onto belief content vary correspondingly. Hence we have only the slenderest of excuses for supposing there to be a single function. Given that the basic notion of a belief-forming function is tied to *a certain way of mapping input features onto belief output content*, and given that many ways are lumped together here in the alleged single complex function, there are strong reasons for positing a large number of perceptual dog-recognition functions, and, in the absence of more compelling reasons on the other side, that is the best choice, apart from the exigencies of epistemological theory.

Hence I am not inclined to worry about the "No-Distinction Problem" in cases like this. To be sure, for any perceptual belief-forming function, no matter how specific, there will be some variation in experiential inputs. Even when I am looking at the maple tree in my front yard from exactly the same distance and angle and at the same time of day, there will still be minor variations from the quality of the light, the condition of the tree, and so on. But we must remember that the effective inputs to a habit, what it "takes account of" in forming the belief, are *abstract* features of the presentation, not that presentation in all its concreteness. Hence the inputs can be the same through considerable variation in the total character of the presentation.

VIII

To sum up. I have suggested that we think of belief formation in a psychologically realistic way, as involving an input-output mechanism (habit)

that yields belief outputs as a certain function of relevant features of inputs. If we do so, we can escape from the dilemma of what choice to make, for epistemic assessment as to reliability, of the type to which a particular belief-forming process belongs. That choice is settled for us by the identity of the function involved in the belief formation in question, for that function is something that possesses a built-in generality. In other words, a particular belief formation is the activation of a general mechanism (habit) that operates in accordance with a certain function. The mechanism (habit) is the psychological realization of that function. On this basis we can say what it is for a belief to be reliably formed.

I. A belief is reliably formed if and only if it was formed by the activation of a reliable belief-forming habit.

And:

II. A belief-forming habit is reliable if and only if it would yield a high proportion of true beliefs in a sufficiently large and varied run of exercises in situations of the sorts we typically encounter.

Notes

- 1 See Alvin Goldman, "What Is Justified Belief?" this vol., ch. 27, and *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- 2 See Marshall Swain, *Reasons and Knowledge* (Ithaca, NY: Cornell University Press, 1981).
- 3 See Frederick Schmitt, *Knowledge and Belief* (London: Routledge, 1992).
- 4 Is this a *conceptual* claim or not? Is it the view that the *concept* of justified belief is the *concept* of a reliably formed belief? Or is the concept of justified belief explained in some other way, reliable belief formation being put forward as a sufficient (and possibly necessary) condition for justified belief as so construed? It could be either. Goldman is quite explicit that reliability functions in the second way for him, whereas in Swain we have what looks like the claim that what we *mean* by saying that a belief is justified is that it is reliably formed (to greatly oversimplify Swain's actual formulation).
- 5 Goldman, "What Is Justified Belief?" p. 613.
- 6 The sequel will illustrate this, at least with respect to Goldman.
- 7 See Richard Feldman, "Reliability and Justification," *The Monist* 68 (1985), pp. 159–74.
- 8 For a valuable discussion of the issues involved here, see Goldman, *Epistemology and Cognition*, sec. 5.5.
- 9 Feldman, "Reliability and Justification," pp. 159–60.
- 10 Alvin Plantinga, "Positive Epistemic Status and Proper Function," *Philosophical Perspectives* 2 (1988), pp. 28–9. See also John Pollock, *Contemporary Theories of Knowledge* (Totowa, N.J.: Rowman and Littlefield, 1987), pp. 116–21.
- 11 Feldman, "Reliability and Justification," pp. 160–1. Plantinga ("Positive Epistemic Status," pp. 29–30) echoes this presentation of Feldman's, and Pollock (*Contemporary Theories*) makes similar claims.
- 12 Plantinga and Pollock give a similar assessment.
- 13 See Feldman, "Reliability and Justification," pp. 168.
- 14 Goldman, "What Is Justified Belief?" pp. 611–12.
- 15 *Ibid.*, p. 614. The reader will note that these two formulations differ with respect to whether it is *truth* or *justification* that is thought of as transferred across the inference. That does not mean that Goldman has contradicted himself. It can be consistently held both that to be "conditionally reliable" a dependent belief-forming process must yield *true* outputs from *true* inputs and that this requirement must be met in

- order for a dependent process to yield a *justified* belief from *justified* inputs. And yet it would seem unwarranted to require this for the justification of an output. That would imply that nondeductive inference never yields justified conclusions from justified premises; for such inference does not always yield true conclusions from true premises. (To be sure, Goldman presents conditional reliability of the process only as a *sufficient* condition for getting justified conclusions from justified premises. But since he mentions no other sufficient condition, one is naturally led to suppose that he takes the condition to be necessary as well.) However, these problems can be ignored for my purposes in this paper.
- 16 A skeptic might question whether we know that such situations are “clearly atypical.” How do we know, she might ask, that we are not always in a Cartesian demon situation? It is no part of my intention here to answer, or otherwise discuss, skepticism. I will only point out that this last qualification to my account, along with all the rest of my suggestions, presupposes the falsity of radical skepticism. My proposals presuppose that we do have some considerable ability to determine when the conditions specified in those proposals are and are not satisfied. It is worth noting that if radical skepticism were correct, everything else would collapse, along with my account of reliability. Skepticism isn’t a problem for it in particular.
- 17 See David M. Armstrong, *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press, 1973), 170.
- 18 In the course of his discussion Armstrong does put certain restrictions on *H*, but none of them have the effect of limiting the extent of the process.
- 19 Goldman, “What Is Justified Belief?” p. 613.
- 20 *Ibid.*, p. 612.
- 21 I am speaking here of a *function* in the mathematical sense, the sense in which, to take a very simple example, addition is a function. Given any two or more numbers, the addition function will yield a unique output as their sum. A function in this sense is, of course, something abstract. That is why I had to specify that a belief-forming process involves the *activation* of a *psychologically realized* function, not just the function as a denizen of logical space.
- 22 The details of this account will vary with one’s favored theory of perception. Since I don’t want to get into those issues here, I am striving for maximum neutrality. Nevertheless, my talk of “presentations” reflects my attachment to a theory of appearing, according to which perceptual experience consists most basically of objects, usually external physical objects, *appearing* to one in certain ways. As a result, sense datum theorists, adverbial theorists, and conceptual-propositional theorists of perceptual experience will not like my way of putting it. I believe, however, that the points I am making here concerning the features of belief formation that are relevant to assessments of reliability are neutral with respect to different accounts of perceptual consciousness. I would invite those who take exception to my formulations to restate what I am saying in their favorite terms.
- 23 Alvin Plantinga, *Warrant: The Current Debate* (New York: Oxford University Press, 1993), 199.
- 24 Goldman, “What Is Justified Belief?” pp. 612–13.
- 25 Goldman, *Epistemology and Cognition*, p. 50.
- 26 Plantinga and Pollock add an extra twist to this formulation. They both hold that the reliabilist needs to find a type that is homogeneous in the sense that there is no subtype of that type the reliability of which differs from the larger type. (See Plantinga, “Positive Epistemic Status and Proper Function,” p. 30, and Pollock, *Contemporary Theories*, pp. 119.)
- 27 See Feldman, “Reliability and Justification,” p. 163.
- 28 *Ibid.*, pp. 164–5.
- 29 He doesn’t claim to have done so.

The Generality Problem for Reliabilism

Earl Conee and Richard Feldman

I Introduction

A Reliabilism and the generality problem

Reliabilism is the most widely discussed contemporary epistemological theory. The most widely discussed version of reliabilism is process reliabilism, which makes the processes that cause and sustain beliefs epistemically crucial. The central idea of process reliability theories of epistemic justification is this:

RJ. A belief is justified if and only if it is produced by a process that reliably leads to true beliefs.¹

A fully articulated reliabilist theory must identify with sufficient clarity the nature of the processes it invokes. In doing so, the theory confronts what has come to be known as "the generality problem."²

A simple example will show the nature of the problem. Suppose that Smith has good vision and is familiar with the visible differences among common species of trees. Smith looks out a house window one sunny afternoon and sees a plainly visible nearby maple tree. She forms the belief that there is a maple tree near the house. Assuming everything else in the example is normal, this belief is justified and Smith knows that there is a maple tree near the house. Process reliabilist theories reach the right verdict about this case only if it is true that the process that caused Smith's belief is reliable. And one might think that the process is

obviously reliable. However, before accepting this conclusion, we should think carefully about exactly what that process is and what its reliability consists in.

Light reflects from the tree and its surroundings into Smith's eyes. Optic neural events result, and these produce further neural events within Smith's brain. Particular concrete occurrences, involving sensory neural stimulation in combination with complex standing conditions in Smith's brain, result in Smith forming the belief. This sequence of concrete events is the process that caused the belief. So, if we take the process that must be reliable to be composed of causally active events that bring about the belief, then reliabilism requires for justification that a sequence of concrete events is reliable.

However, reliability is a kind of tendency. The notion of reliability applies straightforwardly only to enduring mechanisms, such as an eye or a whole visual system, and to repeatable types of processes, such as the type: visually initiated belief formation. Reliability does not apply in any obvious way to the particular sequence of concrete events that caused Smith's belief on this occasion. Each event in the sequence happens only once and the sequence causes whatever beliefs result just on that occasion. Process reliabilists who realize this have sought the requisite reliability in the *types* of process of which particular causal sequences are tokens.³

As many reliabilists have recognized, each token process that causes a particular belief is of numerous different types of widely varying reliability. The token event sequence in our example of seeing the maple tree is an instance of the following types,

Originally published in *Philosophical Studies* 89, 1 (1998), pp. 1-29; reprinted by kind permission of Kluwer Academic Publishers.

among others: visually initiated belief-forming process, process of a retinal image of such-and-such specific characteristics leading to a belief that there is a maple tree nearby, process of relying on a leaf shape to form a tree-classifying judgment, perceptual process of classifying by species a tree located behind a solid obstruction, etc. The number of types is unlimited. They are as numerous as the properties had by the belief-forming process. Thus, process reliability theories confront the question of which type must be reliable for the resulting belief to be justified. It is clear that the answer to this question will significantly affect the implications of the theory. For instance, while visually formed beliefs in general seem to be fairly reliable, processes that use a characteristically maple-leafish visual experience to judge that a maple tree is near seem much more highly reliable, and perceptual processes leading to a belief that a tree, which is behind a solid obstruction, is of a particular species seem generally unreliable, in spite of the fact that in some of their instances, such as the present case, the obstruction is transparent. The process token is of endlessly many other types as well, types of extremely varied reliability. So, which type has to be sufficiently reliable?

Process reliabilists must solve this generality problem. A solution identifies the type whose reliability determines whether a process token yields justification.⁴ This type is "the relevant type" for that token. Thus, it is not the causally active process token that has to be sufficiently reliable, according to reliabilists. It is the relevant type of the process. We need to know what determines this sort of relevance.

Without a specification of the relevant type, process reliabilism is radically incomplete. Only when a bearer of reliability has been identified does the theory have any implications about the justification of beliefs in particular cases. Philosophers often overlook this. They purport to determine whether or not a given belief is justified according to reliabilism using nothing more than one description of the process causing the belief. No such inference is acceptable. The theory must first be elaborated at least enough to imply exactly what process type has to be reliable in the case in question. A fully general reliabilist theory of justification has to do this for all cases in which there is a fact of the matter.

A second necessary task for process reliabilists is to specify which situations of a process type's

operation determine whether or not the type is reliable. Strength of reliability might be settled by the frequency with which the process actually produces true beliefs or rather by its truth-to-falsehood output ratio in certain counterfactual circumstances. The generality problem arises no matter how this question about reliability is answered. William Alston's sensible specification of what determines the reliability of a process type will do for present purposes:

- R. A process type is reliable if and only if it would yield a high proportion of truths over a wide range of situations of the sort we typically encounter.⁵

B Necessary conditions for a solution to the generality problem

A solution to the generality problem must meet the following three conditions.

First, it must be principled. Given the multiplicity of belief-forming process types and their variations in reliability, it is easy to make *ad hoc* case-by-case selections of types that match our intuitions. But case-by-case selections of relevant types does not constitute working out a reliabilist theory of justification.

The claim that the reliability of "the relevant type" of the belief-forming process is what determines a belief's justification is analogous to the claim that "the suitable type" of a horse is what produces victory in a horse race. In the absence of further explanation, this use of "suitable" has no definite content. On its own, the phrase "the suitable type of horse" tells us nothing about what makes horses win races. If there is no further explanation but rather we are offered case-by-case choices of "suitability-making properties," choices made on the basis of knowing which horses are the winners, then the claim is no closer to having any definite content. Clearly, a general basis for identifying suitability is required for the claim to say more than just that something or other makes each winning horse win its race. Analogously, we have an informative reliabilist theory of knowledge or justification only after we are told what determines "the relevant type" in general.

Although a solution must be principled, it need not state necessary and sufficient conditions for relevance that are either precise or always determinate. Claims to the effect that a belief is

“epistemically justified” might be vague and they might be context-sensitive in various ways. A solution must be universal only in that it must specify the relevant type whenever there are definite facts about justification.

The second requirement for solving the generality problem is that the rule must make defensible epistemic classifications. Stating a general rule of relevance that merely assigns some type or other to each process token does not constitute an adequate solution to the generality problem.⁶ The types identified must have a reliability that is plausibly correlated with the justification of the resulting beliefs.

Finally, a solution must remain true to the spirit of the reliabilist approach. We are addressing process reliability theories.⁷ So, the rule of relevance must somehow implement the basic idea that it is the reliability of a process of belief formation, specified in nonepistemic terms, that settles the epistemic status of the belief. Process reliabilists characteristically think that a belief is justified because the workings of the process that produced it (or sustained it) are sufficiently conducive to generating true beliefs. A solution to the generality problem would specify those workings so as to bear out this idea. A solution thus cannot identify the relevant type for a process in a way that merely smuggles a non-reliabilist epistemic evaluation into the characterization of relevant types. For instance, one could develop a form of “reliabilism” that just restates an evidentialist theory of justification in a roundabout way. Pseudo-reliabilism of this sort holds that there are only two relevant types of belief-forming process. One type is “belief based on adequate evidence” and the other type is “belief based on inadequate evidence.” Assuming that the first of these is reliable and the second is not, this version of reliabilism will get plausible results (or at least results that an evidentialist would find plausible).⁸ But this theory is only verbally a version of reliabilism. It mentions the processes of belief formation only in order to characterize the quality of the evidence for the belief. This is obviously incompatible with the spirit of process reliabilism.

C Our thesis

Our thesis is that the prospects for a solution to the generality problem for process reliabilism are worse than bleak. We will investigate the merits

of approaches exemplified by several recent proposals. There is no significant progress in any of these approaches, singly or in combination. The basic process reliabilist idea just does not pan out.

It is reasonable to look for a solution to the generality problem in three places: common sense, science, and context. Common sense is the likeliest source. As we shall soon see, Alvin Goldman’s early account of reliabilism draws much of its initial attraction from the *prima facie* correlation between justified beliefs and beliefs produced by common sense types of processes that are probably reliable. Goldman immediately realized that some refinement of these common sense types is needed, for reasons that we shall illustrate below. But at first glance the thought is appealing that common sense process types like “careful perception,” “vivid memory,” and the like are reliable. So, it makes sense to pursue the reliabilist idea that these types of process produce justified beliefs because of their reliability. In contrast, common sense belief-forming process types like “guessing” seem to be unreliable and seem to yield unjustified beliefs.

If, as we shall argue, common sense types will not do, then the next likeliest source of relevance is scientific classification. Scientific types of belief-forming processes are types that correspond to the predicates that enter into the laws and explanations of science. We shall next investigate the possibility of solving the generality problem by identifying relevant types with these scientific types.

Another reasonable thought is that different types are relevant to justification in different contexts, just as different comparison classes determine the application of terms like “small” and “far” in different contexts and just as different reference classes determine the truth value of probability judgments in different contexts. Thus, we shall consider next the merits of contextualist solutions to the generality problem.

We shall argue that none of these approaches works out. This might raise the concern that our way of posing of the generality problem for reliabilism is somehow ill-conceived. It might be thought that the relevant types are obvious when the question is properly understood, or that no general solution is actually needed. We shall take up this line of thinking as well.

That exhausts the reasonable philosophical approaches to the generality problem. If they all fail, then so does process reliabilism.

II Common Sense Types

In his pioneering defense of process reliabilism, Alvin Goldman appeals to common sense process types in an effort to convey the plausibility of the theory. He writes,

... what kinds of cause confer justifiedness? We can gain some insight into this problem by reviewing some faulty processes of belief-formation, i.e., processes whose belief-outputs would be classed as unjustified. Here are some examples: confused reasoning, wishful thinking, reliance on emotional attachment, mere hunch or guesswork, and hasty generalization. What do these faulty processes have in common? They share the feature of *unreliability*: they tend to produce *error* a large proportion of the time. By contrast, which species of belief-forming (or belief-sustaining) processes are intuitively justification-conferring? They include standard perceptual processes, remembering, good reasoning, and introspection. What these processes seem to have in common is *reliability*.⁹

Thinking of reliabilism in terms of these types gives the theory its initial appeal.

However, common sense types have two liabilities as the basis for a solution to the generality problem. First, there are far too many common sense types to provide a unique identification of the relevant type for each process token. In our initial example, Smith's visually formed maple tree belief results from a process instantiating all the following common sense types: visual process, perceptual process, tree-identifying process, daytime process, indoor process, etc., etc. These types differ widely in their reliability. So, we still need to be told which one determines the justificatory status of the resulting belief.

The other main problem with the types Goldman mentions is that not all beliefs resulting from any one such type are even approximately equally justified. Consider another common sense type that Goldman refers to, brief and hasty scanning. Sometimes, on the basis of a brief and hasty scanning we can get extremely well justified beliefs, as when we see in a glance that there is a tree in the backyard. Other times brief and hasty scanning does not yield a justified belief, as when the belief concerns exactly how many leaves are on the tree. Simple common sense classifications are thus too

broad to make the right epistemic distinctions among beliefs.

In a recent discussion of the generality problem, William Alston sometimes calls what he proposes as the relevant types "habits" of belief formation.¹⁰ Likewise, Charles Wallis appeals to "strategies" of belief formation.¹¹ Habit and strategy are common sense classifications of some of the ways we form beliefs. In classifying trees by species, an expert naturalist has identifying routines that differ considerably from those of novice and ill-informed tree classifiers, even though all of them may judge by experiencing the same views of the trees. The expert is better justified. So there is some initial plausibility in the idea that it is the "routine," the "habit of mind," the "strategy," employed in forming a given belief that determines its level of justification.¹² This suggests:

H. The relevant type for any belief-forming process token is the habit of mind, or belief-forming strategy, that it instantiates.

For a large class of cases, it is doubtful that (H) serves to identify a single relevant type. This is because many process tokens are instances of more than one habit. Smith, our maple tree identifier, may have a habit of concentrating while making careful visual judgments, a habit of calling to mind types of trees known to be in the area when making species classifications, and a habit of counting points on leaves for identifying deciduous trees. Some of her belief-forming process tokens result from the employment of all three habits. So there would be no such thing as "the habit" employed on those occasions, and thus no relevant type by the present proposal.

There are, furthermore, cases in which justified beliefs are formed in a way that is in no intuitive sense "habitual," or "routine," or "strategy-employing." For instance, Smith might happen to notice a cardinal on a branch of the maple tree, and be thereby justified in believing that a cardinal is there. She is not employing any strategy, or habit, or routine, in forming this belief. Thus, a theory that requires a high enough reliability for the relevant type here would conclude that the belief is not justified, since there is no habit or strategy that is either reliable or unreliable. Yet in many such cases the belief clearly is justified.

Also, the same belief-forming habit can produce some justified beliefs and some unjustified ones. Jones might make a habit of judging the theme of a

philosophy article by reading only its concluding paragraph. Sometimes the theme is clearly presented there and Jones will be justified. Other times the final paragraph does not make clear the point of the paper and Jones will not gain justification by employing this procedure.

Another approach using common sense classifications would be to hold that the solution to the generality problem is to classify together processes that produce equally general beliefs:

G. Two process tokens are of the same relevant type if and only if they generate beliefs at the same level of generality.

(G) has no promise as a solution to the generality problem. The problem of finding the relevant type does not reduce to that of finding the right level of generality for the contents of the resulting beliefs. It is often not clear what "level of generality" a belief has. But if there is any merit in the approach that (G) represents, then two judgments will be at the same level of generality if their contents consist in classifying an individual by species. Thus, the following visually based beliefs are all at the same level of generality: this is a mountain-goat, this is a giraffe, this is a crocodile, this is an alligator. (G) implies that all such classificatory beliefs result from the same relevant type, and hence all are equally justified. But clearly this is not so. For instance, some such beliefs are based on more justifying perceptible features than others. To ordinary observers, nearby giraffes are pretty obvious, while nearby crocodiles are easily mistaken for alligators. Processes generating equally general beliefs are not all equally justifying.

A similar idea would be to distinguish processes in terms of the identity of their particular output beliefs, so that the different beliefs just mentioned would result from different relevant types of processes. This has numerous unacceptable results too, however. Clearly there are both justified and unjustified examples of belief in the same proposition.

There is no reason to think that any appeal to simple common sense types will solve the generality problem. Their main liability is that they are too broad to differentiate properly among the justification levels of our various beliefs. Less simple types can be constructed by conjoining together the broad common sense classifications that we have been discussing. These can be much narrower, for instance: visual process causing a belief that classifies by species a close, unobstructed,

opaque object, in bright sunlight. But the members of such types still vary in their degree of justification depending on such things as whether the viewer is familiar with the visual appearance of the species from the viewing angle, has normal vision, is intoxicated, is expecting visual trickery, is emotionally distraught, etc. There is no good reason to believe that even such narrow kinds will include only equally justified beliefs, however elaborately they are specified, as long as they use only common sense nonepistemic categories.¹³

Common sense types thus do not stand scrutiny as candidates to provide a satisfactory solution to the generality problem.

III Science

It is in keeping with the "naturalistic" spirit of reliabilist theories to look for classifying help from natural science. One tempting line of thought is that reliabilists can count on cognitive psychology to identify the types of belief-forming processes that will be useful to their theory. Suggestions of such a view can be found in writings by Alston, Goldman, and Ralph Baergen.¹⁴

A Natural kinds

Alston's mention of habits of mind is not his theoretical proposal for coping with the generality problem. Rather, he suggests that belief-forming process tokens belong to natural kinds and that these kinds are the types to which reliabilists ought appeal. He writes:

With a process token, as with any other particular, any of its properties can be said to be correlated with a type to which it belongs... Even if it is true that you and I belong to indefinitely many classes, such as *objects weighing more than ten pounds*, *objects that exist in the twentieth century*, *objects mentioned in this paper*, etc. etc., it is still the case that membership in the class of human beings is fundamental for what we are in a way that those others are not, just because it is the natural kind to which we belong. I shall suggest that something analogous is true of belief-forming processes – that there are fundamental considerations that mark out, for each such process token, a type that is something like its "natural kind."¹⁵

Although this is not Alston's final account of the matter, it is important to see that more is needed. Merely citing the fact that each belief-forming process falls into a natural kind does not provide an adequate rule of relevance. To see this, note the inadequacy of the following solution to the generality problem.

NS1. The relevant type for any belief-forming process token is the natural kind to which it belongs.

Process tokens may belong to natural kinds. Still, there is no good reason to think that each token belongs to just a single natural kind, and hence no reason to think that (NS1) provides a solution to the generality problem. What the natural kinds of belief-forming processes are is up for grabs, but every belief-forming process token is categorized in multiple ways by laws in each of several sciences. These all seem to be natural kinds of the process, according to current science. Reasonable candidates for natural kinds of a typical visual belief-forming process include electrochemical process, organic process, perceptual process, visual process, and facial-recognition process. All belief-forming process tokens are thus in a multiplicity of natural kinds. So (NS1) does not single out a relevant type for any such process. These natural kinds differ widely in their reliability. So, (NS1) does not solve the generality problem.

B Psychological realism

Process tokens thus belong to numerous natural kinds. Alston contends, however, that for each belief-forming process token there is only one type that is "psychologically real." His suggestion is that this type is the relevant type.

According to Alston, every process token instantiates what he calls a "function." He stipulates that this term is to have its mathematical sense. In the case of beliefs formed on the basis of perceptual experience, these functions take as inputs features of experience to which we are responsive and yield beliefs as outputs. Alston is aware that each particular input/output pair is in the extension of many mathematical functions, but he claims that there is only one such function that any belief-forming process actually is an "activation" of. Only this one is psychologically real.¹⁶

The intended solution to the generality problem seems to be:

NS2. The relevant type for any process token is the natural psychological kind corresponding to the function that is actually operative in the formation of the belief.

(NS2) does narrow the set of candidates for relevant types. Furthermore, psychology does aspire to provide psychological explanations of at least all normally acquired beliefs.¹⁷ If this aspiration is met, there will be psychological types of belief-forming process for all such beliefs.

If (NS2) provides a solution to the generality problem, it must be that there is only one actually operative "psychologically real" type for each belief-forming process. In apparent support of this, while discussing the application of (N2) to beliefs resulting from vision, Alston emphasizes that there is a fact about which elements of a visual scene a person responds to in forming a belief about what is present. Thus, in our example about Smith and the maple tree, Smith might form her belief on the basis of noticing certain features of leaf shape. The token process therefore goes from these input features to that belief. In other examples, when presented with the same scene Smith might pick up on features such as the tree's overall shape or bark texture, rather than leaf shape. These considerations show that the relevant type in the original case must be one that corresponds to a function having as an input/output pair the leaf-shape features to which Smith responds and the belief that she forms.

This may limit somewhat the candidates for relevant types, but in Smith's case there still are numerous overlapping functional relations, and corresponding psychological process types, that include the input/output pair we've identified. There is a very narrow function that goes from just the leaf shape that Smith notices as input to just the output of Smith's particular belief that a maple tree is nearby. There is another function, one that maps a variety of fairly similar inputs, including the particular shape that Smith noticed, onto some belief or other to the effect that there is a maple tree nearby, including the belief Smith forms. There is a broader function, one that maps a variety of somewhat similar inputs, all involving visual shapes, onto either the belief that there is a maple tree nearby or the belief that there is an oak tree nearby or the belief that there is elm tree nearby, etc. There are still broader types that include the original pair, and add new inputs

involving various other sensory cues. In many cases, all these functional causal relations, and many others as well, would be actually operative in forming Smith's belief. Smith's disposition to form the particular belief that she did on the basis of the particular shape that she saw is part of these broader classifying dispositions. The one event of belief-formation manifests them all. Thus, in this and other typical cases, there are a multitude of actually operative psychological types.

An example from another domain may help to make this point clearer. Suppose that a certain pot of water at sea level is brought to a boil. There occurred a certain sequence of concrete events leading to the boiling of the water. This sequence instantiates any number of types, all "physically real." We can identify these types in terms of the functions that describe their final stage. At any given pressure, there is a function that maps water onto a certain temperature – its boiling point. This corresponds to the process "bringing water to a boil at sea-level atmospheric pressure." There is a broader type, "bringing water to a boil." The function corresponding to this second type takes water and varying pressures as inputs, and yields a boiling point for water at each temperature. A still broader function takes as inputs triples of temperatures, pressures, and types of liquid and yields the boiling point for each. This corresponds to the type "bringing liquids to a boil." The token process in our example is an instance of all these types. It is not the case that only one is "physically real." All of them accurately characterize what occurred in the pot. Similarly, far too many functions are "psychologically real." They all correspond to natural psychological kinds. So, (NS2) fails to identify the relevant type.

C Maximum specificity and narrow causal types

Alston also suggests that his psychological realism implies, or at least is compatible with, a different specification of relevant types, one that relies on completely causally specific functions. He assumes that "the functions in question are maximally specific, in that any difference in input that is registered by the function indicates a different function."¹⁸ Making use of this idea of maximal specificity is one way of trying to make good on the idea that only one function is "operative" in the formation of any belief.¹⁹

In any case where a person forms a belief on the basis of a perceptual experience, some features of the experience contribute to a belief-forming causal sequence that starts with the experience. Other features of the experience play no causal role. The same goes for subsequent events in the sequence leading to the belief. Some features of these events help to cause the belief, others do not. The maximum specificity proposal is the idea that the relevant type includes all and only process tokens with the same causal features: they all begin with experiences with the same causally active features, are followed by subsequent events with the same causal features, and have the same belief as output. At one time, Alvin Goldman suggested a very similar solution to the generality problem.²⁰ We can formulate this proposal as follows:

NS3. The relevant type for any belief forming process token *t* is the natural kind that includes all and only those tokens sharing with *t* all the same causally contributory features from the input experience to the resulting belief.²¹

(NS3) does yield a unique type for each process token. But the reliabilist theory of justification that employs (NS3) is seriously defective. (NS3) classifies into the same relevant type only beliefs that share *all* internal causal predecessors. Thus, on the reasonable assumption that the content of any normally formed belief is causally determined by its antecedent psychological causes, according to (NS3) each relevant type can have only one content for its output belief.²² This makes trouble in cases in which the proposition believed dictates the truth-ratio of all process types leading only to it. In such cases the reliability of the relevant type is settled by the mere identity of the belief. Thus, the relevant type of a process leading to any necessary truth must be perfectly reliable. The relevant type of any process leading to any necessary falsehood must be perfectly unreliable. Also perfectly reliable would be the relevant types of all processes leading to any self-confirming belief, such as the belief that someone believes something. The relevant type of the following beliefs would be perfectly unreliable: the belief that there are no beliefs, and the belief that nothing is caused. Since it seems clear that in all of these cases the beliefs can have a level of justification that is other than the implied extreme, these examples run counter to (NS3).

The problems for reliabilist theories built on (NS3) are not confined to beliefs in necessities, impossibilities, or the relatively unusual beliefs just mentioned. Suppose that Jones looks very carefully at a tree and forms the belief that it is a beech on the basis of seeing features which are in fact distinctive to beech trees. As long as experience of such features happens to help to prompt Jones to believe that it is a beech tree, it does not matter to (NS3) why they do so. It can be for good reasons, for bad reasons, or for no reason at all. Recall (R), which tells us that the reliability of a type is determined by the long run truth ratio of its output when it functions under typical conditions. In the normal worlds used to evaluate the reliability of Jones's tree-identifying process, nothing other than a beech tree presents Jones with exactly the features that initiate the causal process leading to his belief.²³ This by itself is enough for the theory to imply that Jones's belief is justified, regardless of how much information he happens to have about the look of beech trees. Since the highly specific causal factors that led to his belief in fact are indicative of only beech trees, his belief must be justified, according to this theory. In the worlds that determine the reliability of the relevant type, only beeches cause the sort of experience that led to his belief that a beech tree is nearby. So this maximally specific type is maximally reliable. Reliabilist theories based on (NS3) thus are unable to distinguish the epistemic status of lucky guesses that happen to be based on distinctive features from expert judgments based on well-understood classifications.

An additional problem is that (NS3) yields a version of reliabilism that is not in keeping with the spirit of process reliabilism. As we have just seen, (NS3) often renders irrelevant the details of the process intervening between an input and a resulting belief. In particular, suppose that Jones and Smith both respond to the same features of a visual input with the belief that there is an elm tree present. Suppose that this input will occur only when there is an elm tree present – it is a distinctive look of an elm leaf, say, the visual appearance of a particular quantity of tiny notches around its edge. Finally, suppose that Smith knows what she is seeing, while Jones is applying some ridiculous and unjustified sort of numerology to the topic. Jones plucks from thin air the idea that the magic number for elms is nine. Jones gets a nine for the tree whose leaf he beholds by counting the number of those distinctive elm notches along the edge of a

leaf, and dividing by six, his "tree number." Given (NS3), the relevant types for their processes are maximally specific. These types are thoroughly reliable since nothing other than an elm would cause just that input in any significant fraction of nearby worlds. The fact that one of the two knows what elms look like and the other does not and the fact that one process goes through a silly application of superstitious nonsense do not affect the reliability of the maximally specific types (NS3) specifies.²⁴ It is just this sort of difference that process reliabilism is supposed to make matter. It is supposed to be sensitive to the possibility that the process one person uses is not generally reliable while the one the other uses is generally reliable, even if in the case at hand both people happen to begin their processes by noticing what is in fact an extremely reliable indicator of the right answer. In other words, process reliability theories are supposed to appeal to much broader relevant types.

D Categories from science

Ralph Baergen discusses several examples, explaining what reliabilists might say about them. By generalizing from his remarks it is possible to devise another way reliabilists might appeal to science to solve the generality problem. It is also a second way of attempting to cash out Alston's remark that only one process type is "actually operative" in belief formation.

One example, discussed in the literature by Richard Feldman, concerns a person who sees something on a distant hill.²⁵ She forms the belief that what she sees is an animal and the belief that it is a mountain-goat. Feldman points out that the more general belief may well be better justified than the more specific one. So, he concludes that reliabilists must find a way to distinguish between the types of processes that cause the beliefs.

Baergen proposes a way to do this.²⁶ He appeals to David Marr's theory of vision, which holds that in classifying objects on the basis of visual perception, we generate a model of the object which "is compared to descriptions in a sort of catalogue. This catalogue is arranged in levels, so that rough categorizations take place at the lower levels, followed by more fine-grained discriminations at higher levels."²⁷ Baergen suggests that we make use of this idea in identifying relevant types:

Our account of processes might well reflect this by saying that rough categorizations are generated by different process[es] than those yielding fine-grained categorizations. Applied to Feldman's case, the mountain-goat belief is generated by a different process than that which generated the animal-belief, for they involve different levels of categorization. Also, the process that generated the animal-belief is likely to be more reliable, for there are likely to be fewer nearby situations in which this generates a false belief than there are for the mountain-goat process. So, Reliabilism *can* provide intuitively correct results here.²⁸

No doubt reliabilists can state a rule of relevance that produces the intuitively correct results "here." But reliabilism needs a fully general rule. Baergen reports part of a theory of vision that implies that perceptual classifications result from processes that are organized by levels of generality of the resulting beliefs. He suggests that reliabilists can identify relevant types in some way that plays on this fact. However, Baergen does not make clear how to build upon this example to develop a general account of relevant types.

One possibility, suggested by Baergen's use of psychology, is that the relevant types are the types that are invoked by the best psychological theories of belief-formation. The idea here is that while any token belongs to numerous types that are psychologically real, only one of those types will enter in the best psychological theory that explains the resulting belief. That type is the relevant type. Marr's theory may have been used to illustrate how this might apply in the case of visual belief formation.

We can formulate this idea as follows:

NS4. The relevant type for any belief-forming process token *t* is the psychological kind that is part of the best psychological explanation of the belief that results from *t*.

It may be that Alston had something like (NS4) in mind when he said that only one type was "actually operative."

(NS4) rests on the dubious assumption that there is a unique "best" psychological explanation for each belief. The value of an explanation depends upon the use to which it is put. A very specific and narrow explanation might have greater value for some purposes, while a broader explanation might have greater value for other purposes.²⁹

Even if (NS4) did identify unique types, it would not be possible to evaluate its implications for process reliabilism without knowing what those types are. There is no good reason to think that the types that are of greatest value for psychological explanation are uniformly helpful to reliabilist theories of justification.

To see why types that are particularly useful for psychological explanation might not be of much help to reliabilists, consider the types Baergen mentions. His proposal ties the relevant types for classificatory beliefs based on visual perception to the level of generality of the resulting belief, and he suggests, plausibly, that a type that produces relatively general beliefs is more reliable than types that produce more specific beliefs. A version of process reliabilism making use of this idea would thus make more general classificatory beliefs better justified than more specific classifications. That is an unacceptable result. Sometimes, a belief applying a broader classification is less well justified than is a belief applying a narrower one. For instance, Jones might use a visual basis for both his belief that the tree he is near is an elm tree and his belief that the tree he is near is a deciduous tree. He can be less well justified in believing the latter, despite its applying a broader classification. This might be true because Jones does not realize that all elms are deciduous and has just a shaky grip on visual cues to deciduous trees, but he has good training in recognizing elms. Similarly, a person may know at a glance that a thing she sees is a whale, but be less well justified in her belief that it is a mammal. Thus, sometimes the more general belief is the more justified, and sometimes not. So the generality of a visually based classificatory belief does not determine a relevant type that yields a satisfactory version of reliabilism. There is, then, no reason to think that the particular scientific classifications Baergen mentions yield types that are entirely helpful to reliabilism.

Although science does provide the tools to narrow the candidates for relevant types, there is no good reason to think that scientific classifications provide the tools for solving the generality problem.

IV Solutions Without a Necessary and Sufficient Condition

Some philosophers have responded to the generality problem by explicitly denying that the problem

requires a general resolution. We will examine two such responses in this section.

A Constraints

Frederick Schmitt proposes five constraints on which process types are relevant, and then appeals to the constraints in describing problem cases.³⁰ According to Schmitt, “relevant processes are cognitive processes.”³¹ His constraints require, among other things, that relevant types are salient, that they are folk psychological process types, and that tokens of the same type are intrinsically similar.

These constraints are not meant to compose what Schmitt calls a “criterion of relevance”: a necessary and sufficient condition for relevant types. Schmitt believes that no such criterion is needed. Instead, the constraints are supposed to identify the sorts of factors that we take to matter when we make judgments about justification.

To explain why no criterion of relevance is needed, Schmitt writes:

[W]e have intuitions about which processes are relevant. In judging whether a subject is justified in an inferential belief, we check to see which inferential process the subject exercises – e.g., whether it is induction from sufficiently many instances or affirming the consequent. We have the intuition that these are the relevant processes to consider. In the case of perceptual belief, we check which environmental conditions obtain – whether it is sunny or foggy – and whether the subject is careful and attentive in perception or quick and distracted. Here again we have intuitions about which processes are relevant. Reliabilism may explain why perceptual or inferential beliefs are justified or unjustified by relying on these intuitions.³²

The existence of these intuitions does not relieve process reliabilists of the responsibility to provide an explanation of their invocation of relevance. Granting that the intuitions exist, the question that we have been asking remains to be answered: According to reliabilism, which type must be reliable for a particular belief to be justified?

Furthermore, Schmitt is mistaken about exactly what intuitions we do have. Schmitt says that “we have intuitions about which processes are relevant.” Since Schmitt is addressing the generality problem, this claim seems intended to imply that

“relevant” in the reliabilist use of “the relevant type of the process” has some intuitive application to examples. But that is not so. The reliabilist use of “the relevant type” is entirely technical. The expression might as well have been “the type that determines justification according to the philosophical theory known as ‘reliabilism’.” No one has pre-analytic intuitions about this topic. It is up to reliability theorists to assign reference to the term from scratch.

Philosophers and others do make intuitive judgments about which features in examples are “relevant” to the justificatory status of beliefs. Schmitt is entirely right to say that in evaluating inferential beliefs we are inclined to judge relevant the pattern of inference followed, and in evaluating perceptual beliefs we judge the environmental conditions and attentiveness of the perceiver to be relevant. We also judge to be relevant the quality and quantity of evidence the believer has. We typically judge to be irrelevant the day of the week on which the belief is formed and the color of the believer’s socks. These are not intuitions about which process types are relevant. They are intuitions directly about what determines a belief’s epistemic justification.

The existence of intuitions about which factors are relevant to justification does not eliminate reliabilism’s need for a theory of relevant types. The constraints Schmitt describes do not do this on their own. They provide a variety of conflicting criteria. In his discussion of cases, Schmitt gives the constraints differing weights so as to achieve the desired result.³³ Perhaps one can, by weighing one factor heavily in one case, a different factor heavily in another, manipulate the constraints in a way that seems to give reliabilism acceptable results. But this is no victory for reliabilism. One could equally well say that the justification of a belief is a function of epistemically irrelevant factors such as the duration of the token of the cognitive process that caused it, the distance of the proximate external cause of the process from the center of the earth, and the amount of energy the process consumed. By *ad hoc* weighting of these factors, one could get acceptable results. The theory, nevertheless, has no merit.

A set of flexible constraints does not solve the generality problem. There are, of course, terms in our language whose application is governed by a set of flexible and varying factors. For example, when we say that someone is a “good athlete,” there are a variety of factors that enter into our

evaluations. They might include speed, strength, and endurance, among other things. But there is no fixed weight uniformly given to these factors. In different contexts these different factors may be weighed differently and it would be a mistake to ask for some fixed ranking of the importance of these various factors in evaluations of athletic ability. Although Schmitt does not say this, it is possible that he intends to propose that evaluations of processes as reliable work in somewhat the same way.³⁴ We turn in the next section to a proposal along these lines.

B Context

Mark Heller contends that the demand for “a general principle for selecting the correct level of generality [for relevant types]... is unreasonable.”³⁵ He thinks that contextual factors determine relevant types and thereby solve the generality problem. Heller elaborates his claims about the role of context as follows:

“Reliable” is a perfectly ordinary word that in perfectly ordinary situations is applied to tokens which are instances of several types, where those types have different degrees of reliability. Yet we somehow manage to use this word without difficulty in ordinary discourse.³⁶

Heller says that the primary task of his paper is to defend the claim that “reliable” is richly sensitive to the evaluator’s context. This much is unobjectionable. The word “reliable” surely is context sensitive. That is, whether or not a thing is accurately called “reliable” depends in part upon the standards set by the context of the ascription. These standards vary, depending for instance on how important it is to rely on the thing that is said to be reliable. This is at most a first step toward solving the generality problem. We need to see how context sensitivity helps with the identification of the relevant type.³⁷

Heller does not claim just that the standards for the application of “reliable” are context dependent. He makes the further claim that we readily understand applications of “reliable” to process tokens that are instances of many types. Thus, when a person says “that process is reliable,” the person can refer to a process token and say something true. The person’s statement is true provided the contextually determined type for that token is

truly said to be “reliable” in the context of attribution. If Heller is right, then context determines two features of our predications of “reliable” to tokens. One has to do with the standard for the strength of reliability required for the term to apply in the context. That feature is of no help in determining the relevant type. The other feature has to do with the identification of the type that must meet those standards. We will refer to these latter types as “contextually determined types.” Thus, a phrase of the form “the process leading to S’s belief that *p*” is supposed to have, relative to a context, a contextually determined type.

A solution to the generality problem can be constructed from these thoughts. The proposal that we shall formulate combines Heller’s contentions about the context dependence of the word “reliable” with the epistemic contextualist view that the standards for assessing the truth value of knowledge and justification attributions is dependent on the attributor’s context.³⁸

C. In any context, *C*, if a person says something of the form “S knows *p*” or “S is justified in believing *p*,” the relevant type of the belief-forming process is the contextually determined type for the phrase “the process leading to S’s belief that *p*” relative to context *C*.

(C) embodies the idea that the description “the process leading to S’s belief that *p*” has a contextually determined process type. (C) puts that idea to the service of reliabilism by identifying contextually determined types with the relevant types needed to fill out reliabilist theories of knowledge and justification.

A fundamental objection to (C) is that contextual factors do not typically yield one determinate process type for the phrase “the process leading to S’s belief that *p*.” As a result, reliabilist theories built upon principle (C) will not yield the correct truth value for many clearly determinate attributions of knowledge or justification.

There are some situations in which phrases referring to process tokens apparently work in the way Heller describes. For example, suppose Jones says, “I have three ways to start my old jalopy: first, shifting into gear while rolling it down a hill; second, jump-starting it; and third, praying and then turning the key. Only the first two usually work.” Suppose that Jones then starts his car by jump-starting it. He remarks:

P. “The process by which I just started my car is reliable.”

Here, Jones's explicit mention of the three types serves to limit drastically the types under consideration. The token mentioned in (P) is of one of those types only. So, this is a case in which "reliable" is explicitly predicated of a process token and we have no problem in understanding what type must be reliable for the predication to be true.

In typical knowledge attributions, however, no contextual narrowing of candidate process types occurs. If it did, then when a person said that someone knows something, there would typically be a range of contextually salient process types such that the token process leading to the person's belief instantiated only one. But this is plainly not the case for most knowledge attributions. Ordinarily, no class of types of belief-forming processes will have been made contextually salient. And nothing else about typical contexts isolates any one type. So, it is just not true that in the context of knowledge attributions there are contextually determined types for the phrase "the process that caused this belief."

To see that this is so, consider our initial example in which Smith comes to know that there is a maple tree nearby by seeing it there. Suppose that Jones, who is sitting in the room with Smith, says:

K. "Smith knows that there is a maple tree nearby."

If Heller's version of reliabilism is to work, there must be, relative to the context of Jones's remark, some contextually determined type for the phrase "the process that caused Smith's belief." What would that type be? Nothing beyond the speaker's intentions seems to narrow the candidate pool in this sort of example. Perhaps Jones would be thinking of something like perception of familiar objects at a reasonable distance, or perhaps to something narrower, such as visual perception of familiar well-lit trees from a reasonable distance. Perhaps Jones would not have any type of belief-forming process in mind. After all, he did not say anything about belief-forming processes and there is no reason to think that he was having any thoughts about them. So, there is no reason to think that in this sort of mundane example, there is such a thing as the contextually determined type for the phrase "the process that caused Smith's belief." Moreover, there is no reason to think that the truth value of Jones's attribution of knowledge to Smith depends in any way on which, if any, of these types Jones has in mind.

Furthermore, even if an attributor of knowledge does have some belief-forming process types in mind, the attributor's thoughts do not identify relevant types in a way that is uniformly helpful to reliabilists. An attributor of knowledge may be mistaken about the reasons for a person's belief, and thus may be thinking about process types that the subject's token process doesn't even exemplify. For example, suppose that Jones witnesses Smith identify a bird as being of a certain species after Smith has had only the briefest glimpse of it under poor lighting conditions. Jones says that Smith's belief is unjustified and so Smith lacks knowledge. Jones does have in mind some process type for Smith's belief, something like forming a bird-classifying belief on the basis of a brief glimpse in poor lighting conditions. Suppose, however, that Smith has formed her belief on the basis of hearing the bird's song, an identification method that Jones has not even thought of. Moreover, Smith does have knowledge as a result. If process reliabilism is anywhere close to the truth about knowledge and justification, it is the reliability of some process type that Smith actually underwent that matters here. So, the generality problem must be solved by appeal to facts about the processes actually involved in the formation of the belief, not by appeal to the possibly mistaken thoughts about those processes in the minds of knowledge attributors.

(C) is incorrect. There simply are no contextually determined types in many, perhaps most, typical contexts in which knowledge and justification claims have a clear truth value. It is true that context helps to determine the standards a process type must meet to be correctly described as "reliable." But the attributor's context comes nowhere near to picking out a relevant type of each belief-forming process, and the process types that are salient to the attributor can be entirely irrelevant to the truth of knowledge claims.

This section has focused on common sense types of belief-forming processes. There are also the many scientific types that classify each belief-forming process. It is clear that nothing about typical contexts of belief, or typical contexts of attribution of knowledge or justification, uniformly singles out one of them. Since our minds are rarely scientifically orientated, speakers' intentions are even less likely to narrow down the scientific types. Nothing else about a context of utterance does so either. Thus, context does not solve the generality problem.

VI Conclusion

That is the full variety of existing approaches to disposing of the generality problem. In the absence

of a brand new idea about relevant types, the problem looks insoluble. Consequently, process reliability theories of justification and knowledge look hopeless.

Notes

- 1 Some authors discuss process reliability accounts of knowledge rather than accounts of epistemic justification. No point will be made below that turns on the differences between knowledge and justification.
- 2 Alvin Goldman in "What is Justified Belief?" this vol., ch. 27, and *Epistemology and Cognition* (Cambridge, MA, Harvard University Press, 1986) defends process reliabilist accounts of epistemic justification. In those works he recognizes the existence of the generality problem. See especially "What is Justified Belief?" pp. 612–13 and *Epistemology and Cognition*, pp. 49–51. The problem is emphasized in Richard Feldman's "Reliability and Justification," *The Monist* 68 (1985), pp. 159–74. It is also discussed by John Pollock in "Reliability and Justified Belief," *Canadian Journal of Philosophy* 14 (1984), pp. 103–14. For responses to the problem, see the works of William Alston, Ralph Baergen, Mark Heller, Frederick Schmitt, Ernest Sosa, and Charles Wallis cited and discussed below.
- 3 It is possible to construct a version of process reliabilism which is only about process tokens and does not confront the generality problem. It faces a considerable problem in making sense of the claim that a token sequence of events has some tendency toward producing beliefs whose truth-ratio would constitute its "reliability." Furthermore, the problems that affect (NS3) below, in virtue of types having just one belief content in their outputs, also affect reliability theories that locate a sort of reliability in process tokens.
- 4 There may not always be a fact of the matter. In the examples used here the belief is either definitely justified or definitely unjustified. The reliability of relevant types for process tokens that lead to beliefs whose epistemic status is unclear will be of less value to present concerns, since such cases are less useful in assessing epistemological theories.
- 5 "How to Think About Reliability," this vol., ch. 28. The proposal mentioned here appears on p. 633. If a satisfactory solution to the generality problem existed, it would be worth addressing difficulties with details of this proposal. For one thing, it is not clear who "we" are supposed to be: all of humanity, or all sentient life on earth, or sentient life everywhere in the universe, or etc. And for another thing, it is unclear which belief-forming situations are "typical." Presumably, bizarre psychology lab situations are atypical. But is perception during space travel atypical, no matter how common it becomes? Are situations of fatigue, intoxication, and excitement atypical? Another difficulty is that we may be specially perceptive during rare emergency conditions. If these are atypical situations, then the justified beliefs from these perceptions might turn out not to be of generally reliable types. In any event, if the present work is correct in its main thesis, then these difficulties are not worth pursuing because the generality problem is insoluble.
- 6 In some passages in "How to Think about Reliability" Alston seems to construe the generality problem somewhat differently. For his purposes, a solution need only show that there are "objective, psychological facts of the matter that pick out a unique type as the one of which a particular process is a token" (p. 628). Thus, he is content to identify relevant types, leaving as a different matter the question of the acceptability of the resulting reliabilist theory. The problem discussed here is that of getting the theory stated *and getting it right*. Any rule of relevance that selects one type for each token will generate some reliabilist theory or other, most of them preposterous.
- 7 Reliabilist theories that make use of the reliability of indicators or mechanisms of belief-formation are thus not our topic. But the problems for the theory of relevance (NS3) below carry over straightforwardly to many reliable indicator theories. Also, there is a problem similar to the generality problem concerning "the mechanism" that produces a given belief. For instance, when a visual judgment relies on only black-and-white discrimination, is the person's whole visual apparatus the relevant mechanism, or is it the black-and-white sensitive portion of that apparatus, or is it only the active part of that portion? Does "the mechanism" for remembered beliefs include parts of the brain active in forming the belief, or just parts active in storing it and recovering it? These questions may have answers that are attractive to reliabilists, but as with the generality problem, the challenge is to identify a principle that implies all and only the correct answers to such questions.
- 8 The results of this theory may be implausible in "demon worlds" in which a demon sees to it that believing in accord with one's evidence does not reliably lead to truths. Whether this is a decisive objection to our evidentialist pseudo-reliabilism depends in part on how reliability is measured. The objection as it is often described makes the chal-

lengeable assumption that a process is reliable in a world only if it regularly leads to truths in that world. In contrast, see for instance William Alston's proposal, stated as (R) above. It does not imply that unreliability in a demon world entails a lack of justification. What (R) makes decisive is roughly the truth-ratio of belief-production in more typical situations.

- 9 "What is Justified Belief?" p. 610.
- 10 "How to Think about Reliability," p. 636ff.
- 11 Charles Wallis, "Truth-Ratios, Process, Task, and Knowledge," *Synthese* 98 (1994), pp. 243–69. See especially p. 266. Wallis relies on belief-forming strategies as part of his response to problems that he discusses for reliability theories of knowledge. It is not clear that he is attempting to solve the generality problem that is the topic of this essay. One reason for this unclarity is that Wallis is working on a concept of knowledge that is relativized to the specification of a task, unlike the traditional concept which is our topic. In any case, we do not intend to attribute to him a simple reliance on strategies as a full solution.
- 12 What follows is a possible solution to the generality problem, suggested by some of Alston's language, that merits a brief look. It is not what Alston proposes. His proposals will be taken up shortly.
- 13 In "What is Justified Belief?" Goldman introduced a distinction between belief-dependent belief-forming processes and belief-independent belief-forming processes. The former processes take beliefs, as well as other factors, as inputs and yield new beliefs as outputs. The latter processes do not take prior beliefs as inputs. Belief-dependent processes are reliable when, over a suitable range of cases, they yield true beliefs if their input beliefs are true. Furthermore, a belief resulting from a belief-dependent process is justified only if the input beliefs are themselves justified. One might hope to appeal to this distinction to help deal with some of the examples discussed in this section and elsewhere in this paper. For example, if one's background beliefs are part of the cause of one's animal-classifying beliefs, then the differences in the degree of justification for the beliefs mentioned here might be attributable to differences in the degree of justification of the beliefs upon which they depend. One might therefore be able to maintain the claim that one relevant type is responsible for all the species-classifying beliefs.

Defenders of reliabilism have not made significant use of the belief-dependent/belief-independent distinction in their efforts to solve the generality problem. There are good reasons for this. First, it is likely that virtually all beliefs that adult humans form are partially caused by other beliefs. Hence, virtually all our beliefs result from belief-dependent processes. It is therefore doubtful that there is any acceptable way for reliabilists to account for the

differing epistemic status of the background beliefs in the examples under discussion. Furthermore, some account of the reliable types for belief-dependent processes is needed. If they are identified in terms of, say, patterns of inference, then process reliabilism turns out to be equivalent to the view that a belief is justified if it results from an inference that is likely to be truth preserving from justified beliefs. This familiar view violates the spirit of process reliabilism since it uses processes only as an indirect way to refer to inferential relations. Finally, it is difficult to see just how to make use of the belief-dependent/belief-independent distinction in conjunction with the specific proposals discussed here. Consider, for example, (G). According to (G), the relevant type is determined by the level of generality of the resulting belief. Thus, according to (G), if two people end up believing that there is a giraffe nearby, they have used processes of the same relevant type. None of the details of the routes by which they got to that belief play any role in determining which type they used. One could be making an invalid inference from justified premises while the other is making an accurate classification based on background knowledge. A theory employing (G) incorrectly evaluates the two beliefs the same way.

As the solutions proposed in the existing literature are discussed below, the reader is invited to note that, like (G), they do not give any role to the difference between belief-dependent and belief-independent processes.

- 14 Alston's and Baergen's implementations of this idea are discussed below. Goldman mentions this sort of approach in *Epistemology and Cognition*, p. 50.
- 15 "How to Think about Reliability," pp. 633–4.
- 16 *Ibid.*, section VI.
- 17 Philosophers often invoke examples in which beliefs result from blows to the head or tumors. It may be that such beliefs do not result from any *psychological* belief-forming process type. Perhaps the explanations of such beliefs must come from a different science or perhaps psychology must be inclusive enough to account for them too, simply because they are mental effects. If some beliefs lack any psychological cause, that would present a problem for (NS2), since even these beliefs can be assessed for justification, and hence they must have a relevant type.
- 18 "How to Think about Reliability," p. 647.
- 19 Throughout this section, when we speak of the maximally specific functions or types, we mean the maximally specific *psychological* functions or types.
- 20 *Epistemology and Cognition*, p. 50.
- 21 Theories can differ over exactly what counts as the input. The process type could begin at the surface of the skin, or farther in at some point where conscious experience begins, or farther out in an external cause of the experience. Alston favors perceptual experi-

- ences as the initial step (pp. 633ff.). He does not defend this selection. No point made here depends on any particular beginning for the causal sequence that constitutes the process.
- 22 Strictly speaking, the assumption may imply only that the “narrow” content of the beliefs resulting from a given relevant type will be the same. No point made here depends on the difference between narrow and broad content. Also, see note 17 above concerning the completeness of psychological explanation.
- 23 It is safe to assume that many of our clear vivid experiences of complex ordinary things like trees are produced only by these same ordinary things in all situations of the sort we typically encounter. Holograms, hallucinations, and perfect pictures are, at most, highly atypical.
- 24 One might think that the fact that Jones relies on unjustified background beliefs has some bearing on this example. That thought seems right. But (NS3) ignores this fact and suggests nothing about how to make use of it in defending a process reliabilist theory. See note 13.
- 25 Feldman, “Reliability and Justification.” The example is discussed on pp. 164f.
- 26 Ralph Baergen, *Contemporary Epistemology*, (Harcourt Brace, Fort Worth, 1995), p. 99. Contrary to what Baergen says, Feldman does not assert that the processes are of the same type. He merely points out the undesirable consequence of the proposition that they are of the same type. It is notable that this sort of example shows that common sense process types, like the visual belief forming process, do not produce beliefs of equal justification even when relativized to a fully detailed specification of the external circumstances.
- 27 *Ibid.*, p. 100.
- 28 *Ibid.*
- 29 Compare the water-boiling example above. There seems to be no reason to think that the explanation at one level of generality is necessarily better than an explanation at any other level.
- 30 *Knowledge and Belief*, (Routledge, Chapman, and Hall, New York, 1992), ch. VI.
- 31 *Ibid.*, p. 169.
- 32 *Ibid.*, pp. 141–2.
- 33 For example, Schmitt says about an example that one constraint, which favors a broad relevant type, outweighs two others that favor a narrower type (*ibid.*, p. 171). In another case, the existence of two constraints favoring a narrower type is said to outweigh one pointing in a different direction (*ibid.*, p. 157).
- 34 Schmitt does say that relevance is a “messy, more contextual affair” than some might think (*ibid.*, p. 159).
- 35 “The Simple Solution to the Problem of Generality,” *Nous* 29 (1995), pp. 501–15. The quotation is from p. 502.
- 36 *Ibid.*
- 37 Ernest Sosa suggests a contextualist response to the generality problem in *Knowledge in Perspective: Selected Essays in Epistemology*, (Cambridge, Cambridge University Press, 1991). Sosa suggests in a programmatic way that relevant types are ones that can “be usefully generalized upon by us as the epistemic community of the” believer (p. 284). Sosa does not elaborate upon this idea, which is a small part of a complex theory. What he does say does not seem to identify a unique type, since multiple types may be “usefully generalized upon.”
- 38 Although the following thesis is suggested by much of what Heller writes, it goes beyond the explicit proposals in Heller’s paper. Also, it makes no use of passages suggesting that a relevant alternatives approach to a theory of knowledge solves the generality problem. We see no plausibility in this latter suggestion on its own, and no way incorporate into it the central theme of Heller’s paper concerning the importance for solving the generality problem of the context-sensitivity of “reliable.”

Externalism and Epistemology Naturalized

Keith Lehrer

Our analysis of complete and undefeated justification in terms of coherence and truth within an acceptance system brings us into conflict with an important competing theory of knowledge called *externalism*. The fundamental doctrine of externalism is that what must be added to true belief to obtain knowledge is the appropriate connection between belief and truth. An earlier account presented by Goldman affirmed that the appropriate connection is causal.¹ This is a very plausible sort of account of perceptual knowledge. The fact that I see something, the hand I hold before me, for example, causes me to believe that I see a hand. The fact that my seeing a hand causes me to believe I see a hand results, it is claimed, in my knowing that I see a hand. According to such an analysis, it is the history of my belief, a matter of external causation, rather than coherence with some internal system, that yields knowledge. The central tenet of externalism is that some relationship to the external world accounting for the truth of our belief suffices to convert true belief to knowledge without our having any idea of that relationship. It is not our conception of how we are related to a fact that yields knowledge but simply our being so related to it.

The early analysis, though providing a plausible account of perceptual knowledge, was a less plausible account of our knowledge of generalities, that men do not become pregnant, for example, or that neutrinos have a zero rest mass, or that there is no largest prime number. For here the nature of the

required causal relationship between what is believed and the belief of it evades explication. That objection is, however, one of detail. Later analyses of others, and of Goldman himself, aim at preserving the thesis of externalism that some relationship of the belief to what makes it true yields knowledge, whether we have any idea of that relationship or not.² Armstrong and Dretske have argued that the relationship should be construed as nomological, one resulting from some law of nature connecting the belief with what makes it true.³ This account is closely connected with the proposal of Nozick that belief track truth in a sense explicated, in part, by the counterfactual claim that the person would not have believed what she did if it were not for the truth of the belief.⁴ Goldman now claims that justified belief must be the result of a belief-forming process that reliably yields truth.⁵ Beliefs resulting from such a process are justified, he contends, while other externalists deny that justification is necessary for knowledge. They all agree, however, that a belief resulting from a certain kind of process or relationship connecting belief with truth can yield knowledge without the sustenance or support of any other beliefs or system of beliefs.

Naturalism

Assuming that the required relationship is something like causation, externalist theories are *naturalistic*. What is a naturalistic theory? It is one in which all the terms used in the analysis are ones that describe phenomena of nature, such as causation, for example, or that can be reduced to such

Originally published in Keith Lehrer, *Theory of Knowledge* (Boulder, CO: Westview Press, 1990), pp. 163–74; reprinted by permission of the publisher, a member of Perseus Books, LLC.

terms. Hume's theory of belief was naturalistic in this sense. He restricted his account of human knowledge to relations of causation, contiguity, and resemblance.⁶ It was, however, Quine who introduced the term *epistemology naturalized* and suggested that inquiry into the nature of human knowledge be restricted to accounts of how belief arises and is altered.⁷ Other philosophers have adopted the term to refer simply to all those accounts of knowledge couched in naturalistic vocabulary or reducible to such a vocabulary. The early account of Goldman considered above according to which S knows that *p* if and only if S's believing that *p* is caused in the appropriate way by the fact that *p* is, in this extended sense, an example of epistemology naturalized. Other early naturalistic accounts offered by Armstrong and Dretske rested on the assumption that the conversion relation was based on nomological rather than causal relations, that is, relations articulated in laws of nature.⁸ Dretske's basic idea was that the reasons we have for believing what we do should be nomologically connected with the truth of what is believed, that is, that it should be a law of nature that a person having such reasons for believing what she does will have a true belief. Assuming a naturalistic account of having a reason which Dretske supplies, such an account is also naturalistic.

One interesting aspect of some externalistic theories which naturalize epistemology is the way in which they attempt to avoid the problems of foundationalism. According to Dretske or Nozick, for example, there is no need either to justify beliefs or posit self-justified beliefs blindly because, contrary to the traditional analysis, the justification of beliefs is not required to convert true beliefs into knowledge. Beliefs or true beliefs having the appropriate sort of naturalistic external relationships to the facts are, as a result of such relationship, converted into knowledge without being justified. It is the way true beliefs are connected to the world that makes them knowledge rather than the way in which we might attempt to justify them. Notice how plausible this seems for perceptual beliefs. It is the way my belief that I see a bird is related to the facts, for example, when my seeing a bird causes the belief that I do, which accounts for my knowing that I see a bird, rather than some justification I have for that belief. What matters for knowledge is how the belief arises, not how I might reason on behalf of it. The traditional analysis says that knowledge is true belief coupled

with the right sort of justification. One sort of externalist analysis says that knowledge is true belief coupled with the right sort of naturalistic relation. It is plausible to assume that the naturalistic relationship will be one concerning how the belief arises, in short, the natural history of the belief. Looked at in this way, the justification requirement can be eliminated altogether in favor of the right sort of historical account.

The Advantages of Externalism

Before turning to details and objections, it is useful to notice the advantages of externalism. First of all, according to some externalists, the need for justification and a theory of justification is eliminated as a component of an analysis of knowledge. On such an account, it is admitted that inference may play some role in the natural history of a true belief, but it is also possible to hold that some beliefs are noninferential. They are beliefs arising from experience without the intervention of inference. This may be offered as an account of what the foundationalist was searching for, but in the wrong place. True beliefs that arise in the appropriate way from experience are knowledge because of the way they arise. There is no need to affirm that such beliefs are self-justified to maintain that they convert to knowledge. We might think of such beliefs as naturalized basic beliefs. Such basic beliefs might, of course, serve as the premises for inferring other beliefs and such inference might convert those beliefs to knowledge as well. It is the history of the belief rather than some sort of justification of the belief that converts it to knowledge.

A reply to skepticism

It is helpful, as well, to notice how neatly this sort of theory deals with traditional and modern forms of skepticism. The skeptic, confronted with a commonsense perceptual claim, that I see a tree, for example, has traditionally raised some skeptical doubt, the Cartesian one, for example, that we might be deceived by an evil demon who supplies us with deceptive sensations which lead us to believe we see external objects when we do not see them at all. Or consider the case of a small object, a "braino," implanted in our brain which, when operated by a computer, provides us with

sensory states which are all produced by the computer influencing the brain rather than by the external objects we believe to exist.⁹ In neither case, affirms the skeptic, do I know I see a tree. The reply is simple. If my beliefs are, indeed, produced by the demon or by the braino, then they are false and I am ignorant. On the other hand, if the beliefs are true and produced in the appropriate way, then I do know.

To this the skeptic is wont to reply that I only know that I see a tree if I know that it is not the demon or the braino that produces my belief and, furthermore, to insist that I do not know this. Why do I not know that there is no demon or braino? I do not know so because my experience would be exactly the same if there were; that is what the demon and braino do, produce exactly the same experiences as I would have if I were to see a tree. I have no evidence whatever against these skeptical hypotheses and, therefore, the skeptic concludes, I do not know them to be false. The reply of the externalist is simple. I do not need to *know* that the skeptical hypotheses are false to know that I see a tree, though, of course, the skeptical hypotheses must *be* false. Otherwise, my belief that I see a tree will be false. All that is necessary is that my belief be true and that it arise in the appropriate way, that it have a suitable history, for knowledge to arise. If my belief is true and has arisen in the appropriate way, then I know that I see a tree, even if I do not know that the conflicting skeptical hypotheses are false. I might never have considered such skeptical machinations. Confronted with them, I might be astounded by them and find them so bizarre as not to be worthy of consideration.

The skeptic might retort that I cannot so easily escape the clutches of skepticism. For example, she might suggest that when I claim to know that I am seeing a car, a Mazda RX7, for example, I must have the information required to tell a Mazda RX7 from cars of another sort, and lacking such information, I do not know that I see a Mazda RX7. Hence, I must know that the car is not a Toyota MR2 or a Porsche 944, which bear some resemblance to a Mazda RX7. Going on, the skeptic might argue that to know that I see a Mazda RX7, I must have the information required to tell seeing a Mazda RX7 from experiences of another sort, those supplied by the demon or braino, and lacking such information, I do not know that I am seeing a Mazda RX7, or even that I am seeing a car. So, the skeptic concludes, just as I must know

that the car I am seeing is not of another manufacture, so I must know that my experiences are not of skeptical manufacture. That, she insists, is precisely what I do not know. Skepticism wins.

Relevant alternatives: a reply to the skeptic

The reply of the externalist is a combination of counterassertion and explanation. The counterassertion is that my true belief that I see a tree arising in the way it does is knowledge, even if I do not know that it has arisen in that way rather than in the way the skeptic suggests. If the skeptical hypothesis is true and the belief has not arisen in the way I suppose, then I lack knowledge, but if it has arisen in the way I suppose, then I have knowledge, even if I do not know competing hypotheses about the origin of the belief to be false. It does not matter whether I know that the belief originated in the appropriate manner. All that matters is that it has originated in that way. Then I know. The explanation about the Mazda, for example, is that there will be some cases, but not all, in which some information excluding other alternatives will be necessary for knowledge. The alternative that I am seeing a Porsche 944 and not a Mazda RX7 is a relevant alternative. The alternative that I am being deceived by an evil demon or a braino is not.¹⁰ What is the difference? My information about what a Mazda RX7 looks like must be sufficient to enable me to distinguish it from other cars, and that information plays a role in the formation of my belief that I am seeing a Mazda RX7. In other cases, particularly those suggested by the skeptic in which there is no such distinguishing information, no such information enters into the appropriate origination of the belief. Where the distinguishing information is a necessary component in the suitable generation of the belief, the alternatives to be distinguished from the truth are relevant, but where it is not a necessary component, the alternatives are not relevant ones. To be sure, a skeptic might find the distinction between relevant and irrelevant alternatives capricious and question-begging as a counterargument. Nevertheless, the initial reply to the skeptic to the effect that true belief originating in the appropriate manner is knowledge, even if we do not know the skeptical hypotheses to be false, is a straightforward consequence of epistemology naturalized whether or not it satisfies the demands of the skeptic.

Knowing That One Knows: Rejection of Deductive Closure

There remains, of course, the question whether I know that I know that I see a tree when I do not know that the skeptical hypotheses are false. If I know that I see a tree, then it follows that the skeptical hypotheses concerning the demon and braino are false. It follows, first of all, from the fact that if I know that I see a tree, then I do see a tree, and, therefore, my experiences are not a result of demonic bewitchment or computer wizardry. It follows, further, from my knowing that I see a tree that my belief originates in the appropriate natural way and not from the demon or braino. In short, it follows both from the fact known and from the knowing of the fact that the skeptical hypotheses are false.

Some naturalists in epistemology would deny that I know that the skeptical hypotheses are false or that I need to know this in order to know that I know that I see a tree. They do this by denying what they call a *deductive closure* condition, namely, the condition that if I know that p and that q is a logical consequence of knowing that p , then I, therefore, know that q . Thus, I might know that p , and know that q is a consequence of knowing that p , even though I do not know that q .¹¹

The denial of closure is directly relevant to replying to the skeptic. I might know that I see a tree, know that the falsity of the demon hypothesis is a consequence of my seeing a tree, even though I do not know that the demon hypothesis is false. If, however, I might know that I see a tree without knowing that the demon hypothesis is false, then might I also know that I know that I see a tree without knowing that the demon hypothesis is false? On the naturalist account, it appears that we may answer in the affirmative. If I can know something without knowing what I know to be the consequences of it, then I can know that I know something without knowing what I know to be the consequences of my knowing it.

The falsity of the demon hypothesis is something I know to be a consequence of my knowing that I see a tree, but I may, nevertheless, know that I see a tree without knowing what I know to be a consequence of my knowing it, to wit, the falsity of the demonic hypothesis. Once we deny the closure condition, we may agree with the skeptic that the falsity of the skeptical hypotheses is a necessary condition of what we know, while cheerfully

admitting that we do not know that the skeptical hypotheses are false. Such are the joys of naturalism and rejection of the closure condition. Given that the appropriate origination of a true belief converts it to knowledge, it becomes obvious that the closure condition must be rejected. My true belief that I see a tree may originate in the appropriate way without a belief in the logical consequences of that true belief originating in the appropriate way. Indeed, I might fail to believe in the truth of the logical consequences. It may strike one as odd that a person should know that she sees a tree, know that the falsity of the skeptical hypothesis is a consequence, and yet fail to know the skeptical hypothesis to be false. The oddity is in the eye of the epistemologist, however, for there is no logical contradiction in this position.

The Naturalistic Relation

The advantages of naturalism are robust, but the theory must be true, not merely advantageous, to solve the problems with which we began. To ascertain whether the theory is true, we must have some account of the naturalistic relationship that is supposed to convert true belief into knowledge. Before proceeding to consideration of such accounts, however, let us consider the rejection of the justification condition. At least one defender of epistemology naturalized, Goldman in his later work, is inclined to argue that the notion of justification is a naturalistic notion. One might be a naturalist about justification and maintain that justification is reducible to some naturalistic relationship. In fact, a philosopher eager to connect the naturalistic analysis with the traditional one might argue that a person has the requisite sort of justification for knowledge if and only if true belief arises in the appropriate naturalistic manner. This would provide us with a naturalistic reduction of justification. Thus, the externalist theory can be construed as a naturalistic account of justification or as a repudiation of a nonnaturalistic account of justification. As we shall see later, however, there are objections to externalist accounts of justification that might lead an externalist to prefer the repudiation strategy.

What exactly is the external relationship that converts true belief into knowledge? It is typical of epistemological theories to take some sort of example as a paradigm of knowledge, to fine-tool

the theory to fit that sort of example and, at least at the outset, to ignore less felicitous examples whose subsequent consideration necessitates rather substantial modification of the theory. That is the history of externalism. The paradigm example is perception. In the case of perception, it is indeed very plausible to contend that what converts perceptual belief into knowledge is the way that the belief arises in perceptual experience. My belief that I see a tree is converted into knowledge by being caused by my actually seeing a tree. Another kind of example is communication. You tell me that Holly Smith is Department Head and that causes me to believe that Holly Smith is Department Head. Do I know that Holly Smith is Department Head as a result of this causation? It might be contended, and has been, that if my informant knows that what he tells me is true, then I know because he knows and his communication caused me to believe this. Of course, his knowing remains to be explicated. The assumption is that there is a causal chain beginning with the fact that Holly Smith is Department Head and ending with my believing it which accounts for my knowing it.

Thus, following Goldman's early proposal, we might consider the following as characteristic of externalistic theories which eliminate the justification condition.

(CK) S knows that p if and only if S believes that p and this belief is caused in the appropriate way by the fact that p .¹²

This account leaves us with the need to explain the difference between being caused in an appropriate way and being caused in a way that is not appropriate. Typical cases of perception provide a model of the appropriate kind of causation.

Dretske has suggested that when x is something S perceives, then

(DK) S knows that x is F if and only if S's belief that x is F is caused or causally sustained by the information that x is F received from the source x by S.¹³

Dretske's analysis, though restricted to perceptual knowledge, highlights two needed qualifications recognized by other authors as well. The first is that the belief need not be caused but only causally sustained by the information that p . This is necessary because the originating causation of a belief might involve an error which is corrected by subsequent information one receives.

If I see two men in the distance, I might take the one on the left to be Buchanan and believe that I see Buchanan when, in fact, it is not Buchanan, as I note when I move closer, but Harnish instead. At the same time, I note that the other man, the one on the right, is Buchanan and that Buchanan and Harnish are dressed in such a way that each appears to be the other in preparation for Tolliver's hallowe'en party. My belief that I see Buchanan was caused by my seeing Harnish dressed as Buchanan, and I continue to hold that belief subsequently when I receive the further information which corrects my mistake about the man on the right but sustains my belief that I see Buchanan and, indeed, that I saw him earlier, though I did not recognize him. Moreover, on this sort of account the appropriate kind of causal relation is explicated in terms of receiving information from a source.

The foregoing analyses are, however, too restricted in scope to provide us with a general analysis of knowledge. There is more to knowledge than perceptual knowledge, and not all knowledge that p can be supposed to be caused by the fact that p . The most obvious example is general knowledge, my knowledge that all human beings die, for example. That fact includes the fact of death of as yet unborn humans which cannot now cause me to believe that all humans die or causally sustain that belief. Our knowledge that all neutrinos have zero rest mass is yet more difficult to account for on such a model, since no one has ever perceived a neutrino at rest. Assuming there to be mathematical knowledge, for example, that integers are infinite, the causal theory seems inappropriate. The integers appear to lie outside the temporal order and to be incapable of causing anything.

Accounts of knowledge in terms of causation or the receipt of information fail to provide an account of our knowledge of general and theoretical truths. Moreover, it is easy to see that externalism in no way requires such a restrictive conception of the external relationship. Causal or information-receiving analyses of knowledge have the virtue of explicating knowledge in a way that explains the connection between truth and belief, between reality and thought, and provides an answer to skepticism. We may, however, maintain the connection between truth and belief without committing ourselves to a restrictive causal connection. Instead, we may require that the *history* of the belief connect the belief with truth.

There are two popular accounts of how the history of a belief might connect the belief with truth. The first and perhaps best known is the later account of Goldman according to which true belief is converted to knowledge *via* justification when the belief is the result of a reliable belief-forming process. Goldman's basic idea, which he has modified and refined, is as follows:

If S's believing that *p* at *t* results from a reliable cognitive belief-forming process (or set of processes), then S's belief in *p* at *t* is justified.¹⁴

The refinements include an account of reliable rules, methods, and processes. The other account, offered by Nozick, requires that a belief must track truth in order to convert to knowledge in the sense that the person would believe that *p* if *p* were true and would not believe that *p* if *p* were not true.¹⁵

The two theories share some advantages. Both retain the reply to the skeptic considered above. They both accomplish this without assuming that we have any guarantee that our beliefs are true, moreover. That my belief is the outcome of a reliable belief-forming process does not presuppose that I have any guarantee of the truth of the belief. Similarly, I might believe that something is true when I would not have believed it, had it not been true even though I have no guarantee that this is so. Thus, given either account of knowledge, the skeptic may be answered while allowing, what seems obvious, that we are fallible in the way in which we form our beliefs, even those converting to knowledge. The result is a fallibilistic epistemology without the postulation of self-justified beliefs.

Objections to Externalism: Information Without Knowledge

There is, however, a general objection to all externalist theories which is as simple to state as it is fundamental. It is that a person who has no idea that her beliefs are caused or causally sustained by a reliable belief-forming process might fail to know because of her ignorance of this. Alternatively, the person who has no idea that she would not have believed what she did had it not been true might fail to know because of her ignorance of that. Any purely externalist account faces the fundamental objection that a person totally ignorant of the external factors connecting her belief with truth, might be ignorant of the

truth of her belief as a result. All externalist theories share a common defect, to wit, that they provide accounts of the possession of information rather than of the attainment of knowledge. The appeal of such theories is their naturalistic character. They assimilate knowledge to other natural causal relationships between objects. Our attainment of knowledge is just one natural relationship between facts among all the rest. It is a relationship of causality, or nomological correlation, or frequency correlation, or counterfactual dependence. But this very attractive feature of such theories is their downfall. The relationship in question may suffice for the recording of information, but if we are ignorant of the relationship, we lack knowledge. As in our refutation of foundationalism, what is missing from the accounts of externalists is the needed supplementation of background information. To convert the specified relationships into knowledge, we need the additional information of the existence of those relationships. Such additional information is, however, precisely the sort of information required for coherence and complete justification.

The general problem with externalism can be seen most graphically by considering the analogy proposed by Armstrong. He suggested that the right model of knowledge is a thermometer.¹⁶ The relationship between the reading on a thermometer and the temperature of the object illustrates the theories mentioned above. Suppose that the thermometer is an accurate one and that it records a temperature of 104 degrees for some oil it is used to measure. We can say, with Armstrong, that there is a nomological connection between the temperature and the thermometer reading, with Dretske that the thermometer receives the information, with Nozick that the thermometer would not record a temperature of 104 degrees if it were not true that the oil was at 104 degrees, and with Goldman that the reading is the outcome of a reliable temperature-recording process. The problem with the analogy is that the thermometer is obviously ignorant of the temperature it records. The question is – why?

One might be inclined to suggest that the thermometer is ignorant of temperature only because it lacks the capacity of thought. If, contrary to fact, the thermometer could entertain the thought that the oil is 104 degrees, would that suffice? Would the thermometer know that the temperature is 104 degrees? What are we to say of this fanciful thought experiment? One might protest, of course,

that it is too far-fetched to turn the philosophical lathe. The thermometer does record information accurately, however, and, given the capacity for thought, it may be said that the thermometer not only contains the information but possesses that information as well. But our thoughtful thermometer does not *know* that the temperature of the oil is 104 degrees as a result of thinking that this is so. The reason is that it might have no idea that it is an accurate temperature-recording device. If it has no idea that this is so, then, even if it thinks the temperature of the oil is 104 degrees when it records that temperature, it has no idea that the recorded temperature is correct. To obtain the benefits of these reflections, however, it is necessary to move to the human case.

Suppose a person, whom we shall name Mr Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system in his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is? Surely not. He has no idea whether he or his thoughts about the temperature are reliable. What he accepts, that the temperature is 104 degrees, is correct, but he does not know that his thought is correct. His thought that the temperature is 104 degrees is correct information, but he does not know this. Though he records the information because of the operations of the tempucomp, he is ignorant of the facts about the tempucomp and about his temperature-telling reliability. Yet, the sort of causal, nomological, statistical, or counterfactual relationships required

by externalism may all be present. Does he know that the temperature is 104 degrees when the thought occurs to him while strolling in Pima Canyon? He has no idea why the thought occurred to him or that such thoughts are almost always correct. He does not, consequently, know that the temperature is 104 degrees when that thought occurs to him.

The preceding example is not presented as a decisive objection against externalism and should not be taken as such. It is possible to place some constraint on relationships or processes converting belief to knowledge to exclude production by the tempucomp. The fundamental difficulty remains, however. It is that more than the possession of correct information is required for knowledge. One must have some way of knowing that the information is correct. Consider another example. Someone informs me that Professor Haller is in my office. Suppose I have no idea whether the person telling me this is trustworthy. Even if the information I receive is correct and I believe what I am told, I do not know that Haller is in my office, because I have no idea of whether the source of my information is trustworthy. The nomological, statistical, or counterfactual relationships or processes may be trustworthy, but I lack this information.

In considering the distinction between belief and acceptance, we might note the argument to the effect that a person who receives the information that p and believes that p as a result may fail to know that p . The reason is that the person may not know that the information she thus receives and believes is correct information. If a person does not know that the information, that p , which she receives is correct information, then she does not know that p . All forms of externalism fail to deal with this problem adequately. To know that the information one possesses is correct, one requires background information about that information. One requires information about whether the received information is trustworthy or not, and lacking such information, one falls short of knowledge. A necessary condition of knowledge is coherence with background information, with an acceptance system, informing us of the trustworthiness of the information we possess.

Externalism and Justification

Some forms of externalism repudiate justification as a condition of knowledge, according to Nozick

and Dretske, for example.¹⁷ Such accounts may provide an interesting account of what it is like for belief to constitute correct information or to track truth, but they provide no account of knowledge. The reason is that no one knows that what she accepts is true when it would have been just as reasonable for her to have accepted the opposite on the basis of her information. A necessary normative condition of a person knowing that p is that it be more reasonable for her to accept that p than to accept the denial of p on the basis of her information. This condition implies the need for a justification condition of the sort we have proposed.

One may, as Goldman illustrates, combine externalism with the affirmation of a justification condition, but such an account, if it takes account of background information in an acceptable manner, will introduce a coherence factor. Goldman insists, for example, that a justified belief resulting from a reliable belief-forming process must not be undermined by other evidence the subject possesses.¹⁸ The condition requiring that the belief not be undermined by other evidence is a kind of negative coherence condition to the effect that the belief not be incoherent with background information. Nevertheless, the source of justification on this account is the reliability of the belief-forming process, that is, the fact that the belief has the sort of history frequently producing true beliefs. As a result of providing a justification condition, a normative constraint is supplied.

The objection raised against externalism in general still applies to such a theory, however. A person totally ignorant of the reliability of the process producing his belief would not know that what he believes is true, even if he had no information that would undermine his belief. The example of Mr Truetemp illustrates this perfectly. He has no evidence that his thoughts about the temperature are incorrect. Had he taken time to consider evidence, he would have discovered that his thoughts about the temperature are correct, but he did not consider any evidence concerning the matter, and that is why he does not know that his thoughts about the temperature are correct.

Take a more commonplace example. If I read a thermometer at the local gas station, and it says that the temperature is 104 degrees, I do not know simply from reading the thermometer that the temperature is 104 degrees. I may not have any evidence that it is untrustworthy, but the competitor to the effect that gas station thermometers are often inaccurate is not one I can beat or neutralize,

at least not without inquiring about the thermometer. Whether or not the belief-forming process is reliable, which perhaps it is, I do not know whether the information about the temperature is trustworthy or not. Indeed, I may have no view on the matter. I may believe what I see out of habit, but this is not knowledge. This is a central problem for externalism, to wit, that ignorance of our reliability or of other external relationships leaves us ignorant of whether our information is trustworthy. Trust sharpens the epistemic blade.

The invincibility objection

There is another objection to historical reliabilism that leads to an important lesson. The objection raised by Cohen is that if we are deceived in such a way that we are invincibly ignorant of the deception, we are justified in what we believe, nonetheless.¹⁹ Cohen's example was the Cartesian demon who deceives us in all our perceptual beliefs. The details of the deception may vary, but let us suppose that the demon clouds our senses and supplies us with deceptive sensory data leading us to believe that we perceive the world though we actually perceive nothing at all. Since our perceptual beliefs are virtually all erroneous, the process that produces them is not reliable. Yet, Cohen suggests, we are certainly justified in our beliefs. We may have done the best we could to ensure that we were not deceived, attended to what we observe with the greatest circumspection, and noticed no error. Having done the best we could, indeed, the best anyone could do, we are certainly justified in believing what we do.

The intuition is reinforced by noting the difference between two people, one who examines his sensory data with the sort of care that would keep him virtually free from error in normal circumstances, and one who forms perceptual beliefs so casually that he would frequently err under the best of circumstances. The former puts together all his information and concludes that he is seeing the path of an alpha particle in a cloud chamber. The other believes this because some person, whom he knows to be scientifically ignorant, has told him that this is what he is seeing. We would wish to say that the former but not the latter was justified in believing that he sees the path of an alpha particle in a cloud chamber, even though both beliefs are produced by processes

that are unreliable, given the interventions of the demon.

Externalism might be modified to meet the objection, and Goldman has suggested more than one way.²⁰ The example shows that it is internal factors, not external ones, that make us justified and explain the difference between the circumspect and casual observers above. The sort of justification appealed to in the example is personal justification as explicated in the last chapter. The circumspect observer wins the justification round arising when the skeptic claims that casual observations are often in error by replying that his observation is circumspect and not casual. The casual observer loses that round to the skeptic.

The absentminded demon

There is, however, an important lesson to be learned from reliabilism. It is that the sort of justification required for knowledge is not entirely an internal matter, either. On the contrary, the needed form of justification depends on the appropriate match between what one accepts about how one is related to the world and what is actually the case. To see this, consider a minor amendment in the preceding example in which the demon, in a moment of cosmic absentmindedness, forgets for a moment to cloud our senses with the result that we really perceive what we think we do. If this moment is one that occurs very briefly as we suddenly awake and is immediately followed by further slumber to conceal the demonic error, we might believe we perceive what, in this instance, we actually do perceive. I might perceive my hand for the first time and believe I see a hand, only to lose consciousness after this formidable event. Do I know that I see a hand in that brief moment? I believe I do, but, since such beliefs are almost all false, I am almost totally untrustworthy in such matters as is everybody else, though accepting myself to be worthy of trust.

I am as much deceived about my trustworthiness in this case as I would be when confronted with a convincing liar who tells me almost all falsehoods about some party he attended except for one fact which, in a moment of absentmindedness, he accurately conveyed, namely, that he arrived before the host. If I accept all that he tells me and also that he is a trustworthy source of information about the event, I may be personally justified in accepting all that he says, but I do not

know that the one truth he has conveyed is a truth. I do not know that he arrived before the host. The reason is that my assumption that my informant is trustworthy is in error, even if he has told me the truth in this one instance, and this error is sufficient to deprive me of the sort of justification I require for knowledge. This is the truth about justification contained in reliabilism.

Complete Justification and Reliabilism

The account that we have offered of complete justification in the last chapter is sufficient to deal with the sort of problem we have just considered. To be personally justified in accepting what another says, one must accept that the person is trustworthy, for, otherwise, the skeptic can win the justification game by claiming that informants are sometimes untrustworthy, or more directly, that the informant from whom I received the information is an untrustworthy informant. Thus, to be personally justified, I must accept that the informant is trustworthy. Since that is false, however, I will not be justified in accepting that my informant arrived before the host on the basis of my verifical system, what is left of my acceptance system when all errors are deleted. I will not be verificaly justified, and so I will not be completely justified either. Hence, the account offered above incorporates the reliabilist insight and explains how we fail to obtain knowledge when the source of information is unreliable.

The appeal of reliabilism and the other forms of externalism may, moreover, be easily understood in terms of the coherence theory and the account of complete justification contained therein. To oversimplify a bit, personal justification depends on our background information about the relationship of acceptance to the truth of what is accepted, about nomological or statistical correlations, about counterfactual dependence, or about reliable processes. This information is contained in my acceptance system. I know that I see my cat sitting on papers on the desk. I accept that I would not believe that I see a cat if it were not true that I see him. I accept that my believing I see a cat is correlated with my seeing a cat, though I would not put it that way. I accept that always, or almost always, I see a cat when I think I see one because my accepting that I see a cat results from a reliable process. It is my acceptance of these things that converts merely accepting that I see a cat into

personal justification, into victory in the justification game. For that victory to be converted into complete justification, however, what I accept about these things must also be true. The conversion of mere acceptance into personal justification depends on my accepting the things about myself whose bare existence the externalist mistakenly assumes to be sufficient to convert true belief into knowledge. The conversion also depends, as the externalist says it does, on these things I accept about myself being true. The error of externalism is to fail to notice that the subject of knowledge must accept that the externalist conditions hold true. The insight of externalism is the claim that the conditions must, indeed, hold true.

Causation and Justification: The Basing Relation

The truth contained in reliabilism is, however, concealed by an error. What a person originally believes as a result of prejudice may later be accepted on the basis of scientific evidence. Therefore, the reliabilist must be in error when he claims that it is what originates a belief that converts it into a justified belief and knowledge. This is, in effect, to confuse the *reason* a person has for believing something with the *cause* of his believing it. The confusion is such a common one that we might name it the *causal fallacy*.

It is easy to see how the fallacy arises. When a person's justification for her belief is based on evidence, then she believes what she does *because* of the evidence. This suggests a causal account of what is involved when the justification of a belief is based on evidence. It suggests that the notion of a justification being based on evidence should be explicated in causal terms. Following this proposal, a person's justification for her belief is based on certain evidence if and only if her belief is causally related in some specified way to the evidence. How to specify the exact way in which the belief must be causally related to the evidence would remain a problem on this approach, but it would be a problem of detail rather than of principle. All such theories must be rejected, however.

Often the evidence on which a justification is based does causally explain the existence of the belief, and it may even be admitted that sometimes the belief is justified because of the way in which it is causally explained by the evidence. Neverthe-

less, it is also possible for a justified belief to be causally independent of the evidence that justifies it. Indeed, it may well be that the evidence in no way explains why the person holds the belief, even though her justification for the belief is based on the evidence. The evidence that justifies a person's belief may be evidence she acquired because she already held the belief, rather than the other way round. This is to be expected, since it is common-sense to distinguish between the reasons that justify a belief and the causes that produce it. The causes of belief are various, and, though the reasons we have for a belief sometimes cause the belief to arise, the belief may also arise from some other cause than having the reasons that justify it. Having the reasons we do may justify the belief, however, even though they have no causal influence upon the belief at all.

An example will illustrate. It is easy to imagine the case of someone who comes to believe something for the wrong reason and, consequently, cannot be said to be justified in his belief, but who, as a result of his belief, uncovers some evidence which completely justifies his belief. Suppose that a man, Mr Raco, is racially prejudiced and, as a result, believes that the members of some race are susceptible to some disease to which members of his race are not susceptible. This belief, we may imagine, is an unshakable conviction. It is so strong a conviction that no evidence to the contrary would weaken his prejudiced conviction, and no evidence in favor would strengthen it. Now imagine that Mr Raco becomes a doctor and begins to study the disease in question. Imagine that he reads all that is known about the disease and discovers that the evidence, which is quite conclusive, confirms his conviction. The scientific evidence shows that only members of the race in question are susceptible to the disease. We may imagine as well that Mr Raco has become a medical expert perfectly capable of understanding the canons of scientific evidence, though, unfortunately, he becomes no less prejudiced as a result of this. Nevertheless, he understands and appreciates the evidence as well as any medical expert and, as a result, has reason for his belief that justifies it. He has discovered that his conviction is confirmed by the scientific evidence. He knows that only members of the other race are susceptible to the disease in question. Yet, the reasons that justify him in this belief do not causally explain the belief. The belief is the result of prejudice, not reason, but it is confirmed by reason which provides the justifica-

tion for the belief. Prejudice gives Mr Raco conviction, but reason gives him justification.

Harman and others, most notably Marshall Swain and Alvin Goldman, have suggested that a belief is based on evidence only if the evidence conditionally or partially explains the belief.²¹ The idea is that, even if the belief is not originated by the evidence on which it is based, it must be causally sustained by the evidence. Again, in the typical case, this will be true. Usually, the reasons a person has for a belief can be expected to have some causal influence on the belief, even if they do not originate that belief. It is, unfortunately, difficult to evaluate the claim that the reasons that justify a belief must always partially explain or causally sustain the belief because a sufficiently precise account of partial explanation and causal sustenance is lacking. There appears to be no better reason for supposing that the evidence that justifies a belief must partially explain or causally sustain the belief than for supposing that it must originate it. The explanation for this is that we may suppose that the evidence justifying Mr. Raco's beliefs does not in any way explain or causally sustain his belief. What explains and sustains his belief is his prejudice. His belief is neither strengthened nor explained by his discovering the evidence for it. His prejudice gives him the strongest level of conviction, and the evidence adds nothing to the strength of it.

One might, however, suggest that his conviction is conditionally or counterfactually explained or sustained by the evidence, nonetheless. It might be proposed that if Mr Raco were not to believe what he does out of prejudice, he would believe it as a result of the evidence. This is again likely, but it need not be so. Imagine that Mr Raco is so dependent on his prejudice that if he were to cease to believe what he does out of prejudice, he would become quite mad and become uninfluenced by reason. To avoid such an objection one might propose, as Swain did, that to say the belief is sustained by the evidence is only to say that if Mr Raco were not to believe what he does out of prejudice but were to continue to believe it nonetheless, then he would believe it as a result of the evidence. Perhaps this is to be expected, but must it be so? Again suppose that were Mr Raco to cease to believe what he does out of prejudice, he would become quite mad and uninfluenced by reason; then were he to believe the same thing though not out of prejudice, he would believe it as a result of madness.

The point is the one with which we began. Though evidence ordinarily has some influence over belief or would have if other factors were to lose their influence, this is really incidental to justification. The analogy between justification and validity explains why. If a person validly deduces a conclusion from something he knows, this may cause him to believe the conclusion or influence his belief in the conclusion. But the validity of the inference does not depend on this causal influence. If valid deduction had no influence whatever on whether a person believed the conclusion, that would not undermine the validity of the inference. Similarly, if someone justifies some conclusion on the basis of something he knows, this may cause him to believe the conclusion or influence his belief in the conclusion. The justification of his conclusion, however, does not depend on the causal influence. Thus, a person may justify a second belief in terms of a first belief and the justification of the second belief may be based on the first without the second belief being causally influenced thereby.

The preceding discussion rests on a distinction between explaining why a person believes something, on the one hand, and explaining how he knows it, on the other. When a person knows that his belief is true, the explanation of why he believes what he does may have something to do with his having the evidence he does, but it need not. The explanation may rest on political, erotic, or other extraneous influences, but the explanation of how a person knows that his belief is true, when the justification of the belief is based on evidence, must be in terms of the evidence. It is how a person knows that is explained by evidence. Why he believes what he does may be explained by anything whatever. Therefore, a justification of a belief that is known to be true is based on certain evidence if and only if his having that evidence explains how he knows that the belief is true. The evidence explains how the person knows, moreover, if and only if the evidence justifies the person's belief. Evidence that justifies a belief consists of that part of the acceptance system of a person which yields complete justification.

The idea of evidence explaining how a person knows may be further clarified by recalling once again that our primary concern is to provide a theory to explain how people know that the information that they possess is correct. If the evidence that a person has justifies her belief that p , then the evidence explains how she knows that

the information that p is correct. She knows this from the evidence. Similarly, if a person is asked how she knows that p , her reply will be to justify the claim that p in terms of her evidence. It is appeal to her evidence that shows that she knows and how she knows. Thus, a justification based on evidence explains how a person knows that p if that justification would be a correct answer to the question "How do you know that p ?"

Reliability and the Justification Game

The example of Mr Raco, a person originally believing out of racial prejudice that the members of some race suffer a disease which members of other races do not suffer and later accepting this on the basis of scientific evidence, shows that a belief need not be produced or, as the example further indicated, even sustained by the evidence that justifies accepting it. Reliability enters into justification not by originating belief but by backing acceptance in the justification game. Consider the justification game played by the prejudiced man before obtaining the scientific information.

Claimant: The members of that race suffer a disease to which members of other races are not susceptible.

Skeptic: You believe what you do as the result of prejudice.

Claimant: It is more reasonable for me to accept that I do not believe what I do as a result of prejudice than to accept that I believe what I do as a result of prejudice. (I am quite unprejudiced concerning members of the race in question, it is just that they are inferior.)

This personal justification would fail to convert into verifical and complete justification. The claimant's error concerning his prejudice would disqualify this move in the verifical justification game.

After acquiring the scientific information, the claimant is in a position to neutralize the claim of the skeptic in the justification game by making the following reply to the claim of the skeptic above:

Claimant: It is as reasonable for me to accept that I believe what I do out of prejudice and that the best scientific evidence shows that what I thus believe is, in fact, true than to accept merely that I believe what I do out of

prejudice. (In the standard medical reference work concerning this disease, it is stated that only members of the race in question are susceptible to the disease. This has been confirmed by recent studies cited in . . .)

This move succeeds in the verifical justification game. The claimant wins the round, and his move cannot be disqualified. Whatever his moral failings, as a result of obtaining scientific understanding, he is victorious in the justification game. He is, therefore, personally and verifically justified in accepting what he does.

The preceding reflections illustrate the point that the evidence which justifies a person in accepting something must explain how the person knows that p rather than why he believes it. The scientific evidence explains how the person knows by explaining how he is victorious in the justification game. Usually, what makes a person victorious in the justification game is closely connected to what makes him believe what he does. But the connection is not essential to justification. As a result, the reliability essential to justification is not the reliability of the process which produces or causally sustains belief. What is essential is the reliability or trustworthiness of the evidence for what we accept to guide us to acceptance of what is true rather than false. The trustworthiness of the evidence makes us trustworthy in the matter, whatever our general defects. In epistemology as in life generally, you do not have to be perfect in order to be justified.

Externalism, Foundationalism, and Coherence: An Ecumenical Reconsideration

The foregoing articulation of the coherence theory of justification suggests that there is some merit in the foundation theory and in externalism which we have preserved in our theory. It is, therefore, time to turn from criticism to ecumenicalism. The foundation theory held some introspective, perceptual, and memory beliefs to be self-justified. We argued that the justification of all such beliefs depends on background information concerning our trustworthiness in such matters. Thus, it is coherence with such information in our acceptance system that produces the justification. Nevertheless, we concede that some beliefs are justified without inference because we accept ourselves to

be trustworthy in such matters, and that a principle of our trustworthiness is needed to convert mere acceptance into justified acceptance.

Moreover, though the principle of our trustworthiness must cohere with what we accept about our successes and failures in past epistemic employments, the principle of our own trustworthiness provides its own personal justification. We are, at least in part, personally justified in accepting that we are trustworthy precisely because we accept that we are. If we did not accept that we were trustworthy, there would be an unbeatable skeptical challenge to any claim we made in the justification game, to wit, that we are untrustworthy in what we accept. To beat that move, we must accept that we are trustworthy. So, there appears to be at least one thing that we accept, one important and fundamental thing, that is self-justified as the foundationalist contended, even if it is not those introspective, perceptual, and memory beliefs that he most favors. To be personally justified one must accept some principle of trustworthiness which is, in part, self-justified.

To be verifiably and completely justified as well, some principle of trustworthiness we accept must be true. Otherwise, the skeptical challenge that we are not trustworthy in what we accept would not be beaten in the verifiability justification game. The insight of externalism is the contention that there must be some truth connection between our accepting something and the truth of what we accept. It is our acceptance of our trustworthiness and the correctness of what we thus accept that yields the truth connection.

Externalism is motivated by the doubt about whether what we accept can supply the truth connection. The reason for the doubt is the assumption that it is psychologically unrealistic to suppose that beliefs about our beliefs are necessary for knowledge. Such higher order beliefs about beliefs are not, of course, necessary for receiving and relaying information. Even a thermometer is capable of that. Such beliefs are, however, necessary for knowledge. Is it unrealistic to suppose that people believe themselves to be trustworthy? Some unrealistic theory of belief may yield that consequence, but our theory of acceptance avoids it. The mental state of acceptance is a functional state, one that plays a role in thought, inference, and action. We think, infer, and act in a way manifesting our trust in what we accept.

Thus, it is appropriate and not at all unrealistic to suppose that, in addition to the other things we accept, we accept our own trustworthiness as well. We have supplied the truth that supplies the truth connection required by the externalist in the form of a self-justified principle of our own trustworthiness. We cannot be accused of chauvinism in claiming that complete justification is the result of coherence with an acceptance system incorporating the principle. Unless we are trustworthy in what we accept, neither we nor our adversaries can be justified in what we accept and we must all concede the day to the skeptic. If we are trustworthy, as we accept ourselves to be, what we accept will cohere with our acceptance system and our verifiability system to yield complete justification. The attainment of knowledge, like so many other benefits in life, rests on self-trust.

Notes

- 1 Alvin Goldman, "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1987), pp. 357-72.
- 2 Alvin Goldman, *Epistemology and Cognition* (Cambridge: Harvard University Press, 1986).
- 3 D. M. Armstrong, *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press, 1973); Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981).
- 4 Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981), ch. 3.
- 5 Alvin Goldman, "What Is Justified Belief?" this vol., ch. 27; see also his *Epistemology and Cognition*.
- 6 David Hume, *A Treatise of Human Nature* (Indianapolis: Bobbs-Merrill, 1977).
- 7 W. V. O. Quine, "Epistemology Naturalized," this vol., ch. 23.
- 8 Armstrong, *Belief, Truth, and Knowledge*; Fred Dretske, *Seeing and Knowing* (London: Routledge and Kegan Paul, 1969).
- 9 The "brain" example comes from James M. Cornman, Keith Lehrer, and G. S. Pappas (eds), *Philosophical Problems and Arguments: An Introduction* (New York: Hackett, 1987), 3rd edn, pp. 54-5.
- 10 Dretske and Goldman have both discussed replies to skepticism that distinguish relevant from irrelevant alternatives. See Goldman, "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976), pp. 771-91, and Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971),

- pp. 1–22. See also Dretske, *Knowledge and the Flow of Information*.
- 11 See Nozick, *Philosophical Explanations*, ch. 3, and Dretske, *Knowledge and the Flow of Information*.
- 12 Goldman, “A Causal Theory of Knowing.”
- 13 Dretske, *Knowledge and the Flow of Information*.
- 14 Goldman, “What Is Justified Belief?”
- 15 Nozick, *Philosophical Explanations*.
- 16 Armstrong, *Belief, Truth, and Knowledge*.
- 17 Nozick, *Philosophical Explanations*; Dretske, *Knowledge and the Flow of Information*.
- 18 Goldman, *Epistemology and Cognition*, pp. 62–3, 111–12.
- 19 Stewart Cohen, “Justification and Truth,” *Philosophical Studies* 46 (1984), pp. 279–96; see also Keith Lehrer and Stewart Cohen, “Justification, Truth, and Coherence,” *Synthese* 55 (1983), pp. 191–207.
- 20 Goldman, *Epistemology and Cognition*, pp. 107–9, and “Strong and Weak Justification,” in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 2 (Atascadero, CA: Ridgeview, 1988), pp. 51–70.
- 21 See Gilbert Harman, “Knowledge, Reasons, and Causes,” *Journal of Philosophy* 67 (1970), pp. 844–55; also his *Change in View* (Cambridge: MIT Press, 1986) and *Thought* (Princeton: Princeton University Press, 1973); and Marshall Swain, *Reason and Knowledge* (Ithaca: Cornell University Press, 1981). For an argument that no revision of Swain’s account can succeed, see Jonathan Kvanvig. “Swain on the Basing Relation,” *Analysis* 45 (1985), pp. 153–8.

Externalism and Skepticism

Richard Fumerton

After examining a number of different controversies associated with the internalism/externalism debates in epistemology, I argued that two of the most fundamental issues separating internalists and externalists are the question of whether fundamental epistemic concepts can be "naturalized" and the question of whether one takes access to inferential connections to be a necessary condition for inferential justification. We introduced the labels "inferential internalism" and "inferential externalism" to refer to the two positions one might take on this last question.

In this chapter I am primarily interested in exploring the ways in which an externalist might respond to the classic skeptical arguments sketched previously. For convenience I focus primarily on reliabilism, but almost all of what I say will apply *mutatis mutandis* to other paradigmatic externalists. My first aim is simply to be clear about the framework within which a foundationalist version of externalism will face the skeptical challenge. I want to understand where externalism leaves the philosopher when it comes to approaching normative epistemological issues in general, and these issues as they relate to skepticism in particular. But, as I implied earlier, I think the very examination of the way in which the philosophical externalist should approach skepticism may reveal the fundamental weakness of externalism as a metacpistemological account of concepts fundamental to philosophical concern with epistemology.

Externalism, Foundationalism, and the Traditional Skeptical Argument

Previously, I have tried to characterize what I take to be the fundamental structure of skeptical arguments. The skepticism I am most interested in is skepticism with respect to justified or rational belief. Furthermore, we are concerned, in the first place, with "local" rather than "global" skepticism. The skeptics we considered put forth arguments designed to establish that we have no justified beliefs with respect to certain *classes* of propositions. They offered skeptical arguments that concluded that we have no reason to believe propositions about the physical world, the past, other minds, the future, and so on. The traditional skeptic virtually always presupposed some version of foundationalism, presupposed that we do have noninferentially justified belief in at least some propositions. The presupposition was seldom stated explicitly, but one cannot read any of the important historical figures concerned with either advancing or refuting skepticism without reaching the conclusion that they took some propositions to be epistemically unproblematic, where their unproblematic character seemed to stem from the fact that one did not need to *infer* their truth from any other propositions believed. In both the rationalist and the empiricist tradition, at least some propositions about the content of one's current mental states were taken to have this unproblematic, noninferential character.

The first step, then, in advancing an argument for skepticism with respect to some kind of proposition is to establish that our access to the relevant truth is at best *indirect*. In the terminology we

have developed, the skeptic begins by denying that we have noninferential knowledge, or noninferentially justified belief in the relevant sort of proposition. Thus, for example, skeptics with respect to the physical world deny that we have noninferential "direct" access to physical objects. The standard claim is that if we have justification for believing anything about the physical world, that justification reduces to what we can legitimately *infer* about the physical world from what we know about the character of our past and present sensations. The skeptic about the past claims that we have no direct – that is, noninferential – knowledge of the past. What we know or reasonably believe about the past is restricted to what we can legitimately infer about past events from what we know about the present state of our minds.

One of the primary advantages that paradigmatic externalist accounts have in the battle against skepticism is the ease with which they can deny the crucial first premise of skeptical arguments. The class of noninferentially justified beliefs is likely to be much larger given an externalist epistemology. Notice that I say "likely" to be much larger. As far as I can see, virtually all externalist epistemologies entail that it is a purely contingent question as to which beliefs are justified noninferentially and which are not. On the reliabilist's view, for example, the question of whether or not one is noninferentially justified in believing at least some propositions about the physical world is a question about the nature of the processes that yield beliefs about the physical world and the nature of their "input." If we have been programmed through evolution to *react* to sensory stimuli with certain representations of the world, and we have been lucky enough to have "effective" programming, then we will have noninferentially justified beliefs about the physical world. If Nozick is right and our beliefs track facts about the physical world around us, and this tracking does not involve inference from other propositions, we will again have noninferentially justified beliefs about the physical world. If our beliefs about the physical world are acting like that reliable thermometer that Armstrong uses as his model for direct knowledge, if we are accurately registering the physical world around us with the appropriate representations, then again we have noninferential, direct knowledge of that world. Whether or not we have such noninferential justification for believing propositions describing the physical world, on any of these

externalist ways of understanding noninferential justification, is a purely *contingent* matter.

That it is a contingent fact is not in itself surprising, nor is it a consequence peculiar to externalist epistemologies. It is certainly a contingent fact on the acquaintance theory that I am acquainted with the fact that I am in pain. It is a contingent fact that I am in pain and so obviously contingent that I am acquainted with it. It is less obvious on traditional foundationalisms that it is a contingent fact that we are *not* acquainted with certain facts. It might seem, for example, that one *could not* be acquainted with facts about the distant past, the future, or even the physical world if it is understood as a construct out of actual or possible experience or as the cause of certain actual and possible experience.¹ But even here it is difficult to claim that it is necessarily the case that conscious minds are not acquainted with such facts. There may be no God, but it is not obvious that the concept of a consciousness far greater than ours is unintelligible. If the concept of a specious present makes sense, such a consciousness may have the capacity to directly apprehend a much greater expanse of time than can finite minds. In any event, it is not *clear* that the class of facts with which *we* can be acquainted exhausts the facts with which all possible consciousness can be acquainted.

But even if the scope of noninferentially justified belief is contingent on both internalist and externalist versions of foundationalism, there are crucial differences. On traditional (internalist) versions of foundationalism, philosophers are at least in a position to address reasonably the question of the content of noninferentially justified belief. The philosopher is *competent*, at least as competent as anyone else, to address the questions of whether or not we have noninferentially justified beliefs in propositions about the physical world, for example. There are two sources of knowledge as to what we are noninferentially justified in believing. One is dialectical argument. The other is acquaintance itself. One can be directly acquainted with the fact that one is directly acquainted with certain facts.

On the classic externalist views, the facts that determine whether one is noninferentially justified in believing a proposition are complex nomological facts. Given paradigm externalism, it is not clear that a philosopher *qua* philosopher is even in a position to speculate intelligently on the question of whether or not we have noninferentially justified belief in any of the propositions under skept-

tical attack.² Because the externalist has reduced the question of what is noninferentially justified to questions about the nature of the causal interaction between stimuli and response, and particularly to the processes of the brain that operate on the stimuli so as to produce the response, the search for noninferential justification would seem to be as much in the purview of the neurophysiologist as the philosopher.³ In the last two hundred years, the vast majority of philosophers simply have not had the training to do a decent job of investigating the hardware and software of the brain. But without this training, it hardly seems reasonable for philosophers to be speculating as to what is or is not a reliable belief-independent process. To be sure, some contemporary epistemologists are trying to "catch up" on developments in cognitive science and even neurophysiology, but I cannot help worrying that the experts in such fields will quite correctly regard these philosophers as simply dilettantes who, having tired of their a priori discipline, now want to get their hands dirty in the real-life work of science.

Given this possibility, it is ironic that so many philosophers find externalist analyses of epistemic concepts attractive precisely because they seem to capture the prephilosophical intuition that there is something direct about our knowledge of the physical world through sensations. Through sheer repetition of the arguments, many philosophers got used to talking about *inferring* the existence of a table from propositions about the character of sensation, or *inferring* propositions about the past from propositions describing present consciousness. But critics have correctly pointed out that if such claims are intended to be phenomenologically accurate descriptions of our epistemic relation to the world, they are hardly credible. Anyone who has tried to draw knows that it is very difficult to distinguish the world as it appears from the world as it is. That there is a conceptual distinction between phenomenological appearance and reality seems obvious. If the difficulty of artistic representation shows that we rarely reflect on appearances (as opposed to reality), it also shows that there is such a thing as appearance. A number of philosophers have argued that the most frequent use of "appears" terminology is not that of describing the phenomenological character of sensation, but rather that of expressing tentative belief.⁴ When I say that he appears to be a doctor I am probably only indicating my tentative conclusion that he is a doctor.

But even if we recognize what Chisholm called the "epistemic" use of "appears," there is surely another use of the term that is designed to capture the intrinsic character of sensation. When I say that the people on the street below look like ants, I am not expressing the tentative conclusion that they are ants. Again, as Sellars pointed out, we cannot directly conclude from such examples that the descriptive use of "appears" gives us a "pure" description of experience uncontaminated by reference to the physical world. "Appears" sometimes has what Chisholm called a "comparative" use.⁵ To say that *X* appears *F* in this sense is to say that *X* appears the way *F* things appear under some set of conditions. The people down below look like ants in the sense that they look something like the way ants look when you are relatively close to them under standard conditions. Such complex facts include reference to physical objects and their tendency to appear in certain ways under certain conditions, and consequently are implausible candidates for objects of direct acquaintance. But Sellars aside, it is difficult to avoid the conclusion that the comparative use of "appears" virtually presupposes some other way of understanding the phenomenological character of appearance. There is some way that things appear and it is that way of appearing that the artist must think about in trying to represent realistically some aspect of the world. But whether or not this "noncomparative"⁶ use of appears exists and is intelligible, it does not alter the phenomenological fact that we do seldom, if ever, consciously infer propositions about the physical world from propositions describing the character of sensation.

We also seldom consciously infer propositions about the past from anything we might call a memory "experience." The very existence of memory "experience" is far from obvious. And it is relatively seldom that our commonplace expectations about the future are formed as a result of careful consideration of premises describing past correlations of properties or states of affairs. When I expect my next drink of water to quench my thirst instead of killing me, I do *not* first consider past instances of water quenching thirst. It is useful to reflect carefully on this fact, for even most externalists will view this kind of knowledge as involving inductive *inference*. We must, therefore, be cautious in reaching conclusions about the role of phenomenology in determining whether a *justification* is inferential or not. We must distinguish questions about the causal origin of a belief

from questions about the justification available for a belief.

We must also distinguish between occurrent and dispositional belief. It may be that I have all sorts of dispositional beliefs that are causally sustaining my beliefs when I am completely unaware of the causal role these dispositional beliefs play. In introducing this discussion I suggested that it was ironic that externalists would find attractive the fact that their externalism can accommodate the apparent phenomenological fact that far fewer commonsense beliefs involve inference than are postulated by traditional foundationalism. The irony is that phenomenology should have no particular role to play for the externalist in reaching conclusions about what is or is not inferentially justified. According to the externalist, the epistemic status of a belief is a function of the nomological relations that belief has to various features of the world. These nomological facts are complex and are typically not the kinds of facts that have traditionally been thought to be under the purview of phenomenology. I suppose an externalist can define some belief-producing process as "phenomenological." But again, even if one can describe such a process, it will be a contingent question as to what beliefs such a process might justify, a contingent question that goes far beyond the competency of most philosophers (and certainly most phenomenologists) to answer.

But perhaps I am being unfair in suggesting that the philosopher who is an externalist in epistemology has no particular *credentials* qualifying him to assess the question of whether the skeptic is right or wrong in denying the availability of noninferential justification for beliefs under skeptical attack. The skeptics, after all, had arguments in support of their conclusion that we have no noninferentially justified beliefs in propositions about the physical world, the past, the future, other minds, and so on. The externalists can at least refute those arguments based on their a priori reasonings about the correct metaepistemological position. The most common way of supporting the conclusion that we do not have noninferentially justified beliefs about the physical world is to point out that we can imagine someone having the very best justification possible for believing that there is a table, say, before him, when the table is not in fact there. A person who is vividly hallucinating a table can have just as good reason to think that the table exists as you do. But we can easily suppose that there is no table present before

the victim of hallucination. If *direct* epistemic access to the table is anything like a real relation, then it cannot be present when the table is not present. But if the victim of hallucination does not have direct access to the table, and the victim of hallucination has the same kind of justification you have for thinking that the table exists (when you take yourself to be standing before a table in broad daylight), then you do not have direct access to the table either.

The reliabilist will deny the association between noninferential justification and direct access to the table. To have a noninferentially justified belief about the table's existence is to have a belief about the table produced by an unconditionally reliable belief-independent process. The victim of hallucination has (or at least might have) a belief in the table's existence produced by an unconditionally reliable belief-independent process. It depends in part on how we define the relevant process. But if we think of the stimuli as something like sensations (which the hallucinator has), and the process as what goes on in the brain when sensation is assimilated and turned into representation, there is no reason why someone who is hallucinating cannot satisfy the conditions for having a noninferentially justified belief, assuming of course that the process in question really is unconditionally reliable. The reliabilist's metaepistemology allows at least a conditional response to the skeptic's attack. More precisely, the reliabilist can point out that a reliabilist metaepistemology entails that the skeptic's conclusion about the noninferential character of belief about the physical world does not follow. And, of course, everything the reliabilist says about the physical world applies to the past, other minds, and even the future. The reliabilist probably will not claim that beliefs about the future are noninferentially justified, but he should claim that there is no reason in principle why they could not be, and should continue to assert that the skeptic has no argument for the conclusion that we have no direct, that is, noninferentially justified, beliefs about the future.

Interestingly, not all externalists will reject the skeptic's claim about noninferential justification in the same way. Consider again the reliabilist's response to the argument from hallucination as a way of establishing that we have no noninferentially justified beliefs about the physical world. The crucial move for the reliabilist was to deny that we are forced to regard the hallucinatory situation as one in which the subject lacked a

noninferentially justified belief. A causal theorist about direct knowledge, like Armstrong, might admit that in hallucinatory experience we lack noninferential knowledge, but continue to assert that in veridical experience we have such knowledge. This externalist is more likely to deny the skeptic's presupposition that we should say the same thing about the nature of the justification available to the victim of vivid hallucination and the person who has *qualitatively* indistinguishable veridical experiences. You will recall that there is one sense of "internalism" according to which the internalist holds that the conditions sufficient for justification are always states internal to the subject. If sensations are not themselves relations (a controversial claim, to be sure), and the sensory evidence of *S* and *R* is indistinguishable, and there is nothing else "inside their minds" to distinguish their epistemic state, then this internalist will insist that if the one has a certain kind of justification for believing something, then so does the other.

But a causal theorist thinks that the relevant question that determines the nature of the justification available for a belief involves the *origin* of the belief. The internal, that is, nonrelational, states of *S* and *R* can be qualitatively indistinguishable, but *S*'s internal states can result in *S*'s having a noninferentially justified belief by virtue of their being *produced* in the appropriate way. *R*'s internal states might bring about the very same belief, but because they were not caused by the appropriate facts they will not result in the having of a noninferentially justified belief. In short, the hallucinator's belief cannot be traced via sensation back to the fact about the world that would make the belief true. The person lucky enough to have veridical experience typically has a belief that can be traced back to the fact that makes the belief true. This is a perfectly clear distinction, and there is nothing to prevent an epistemologist from arguing that this just is the distinction that determines whether or not someone has a justified or rational belief. Furthermore, the question of whether the justification is inferential has only to do with the kinds of links in the causal chain leading to the relevant belief. If the causal connection goes directly from some fact about the physical world, to the occurrence of sensory states, to representations about the physical world, then there are no other *beliefs* that crucially enter the story. The justification that results will be justification that does not logically depend on the having

of other justified beliefs.⁷ It will be noninferential justification. So again, we can see how an externalist metaepistemology can put one in a position to claim that the skeptic has not established the crucial premise concerning the inferential character of our belief in the propositions under skeptical attack.

Even if externalism allows one to point out that the skeptic has not established the crucial first premise of the argument, it does not follow, of course, that the externalist has given any positive reason to suppose that the skeptic is wrong in claiming that the propositions under skeptical attack are not the objects of noninferentially justified belief. Both skeptics and nonskeptics play on a level playing field. There is no "burden of proof" when it comes to fundamental issues in epistemology. If the philosopher wants to claim that we have noninferentially justified belief in certain propositions, then the philosopher can give us good reasons to think that such justification exists. The skeptic who wishes to deny that we have such justification can give us good reasons to think that it does not exist. The skeptic, however, also has a fall-back position. Without arguing that we have no noninferentially justified beliefs in propositions about the physical world, the past, other minds, and the future, the skeptic can move "up" a level and deny that we have any good reason to believe that we have noninferential justification for these beliefs. A strong access internalist can move from the proposition that we have no justification for believing that we have a noninferentially justified belief that *P* to the conclusion that we do not have a noninferentially justified belief that *P*. But the externalist rejects just such an inference. Even if we abandon strong access internalism, however, we might find skepticism that maintains that we have no justification for believing that we have a justified belief that *P* just as threatening as skepticism that concludes that we are unjustified in believing *P*. Before we consider the question of whether skepticism will arise at the next level up within an externalist epistemology, let us briefly discuss the externalist approach to normative issues involving inferential justification.

Skepticism, Externalism, and Inferential Justification

Most of the general observations made about the externalist's response to skeptical challenges con-

cerning the class of noninferentially justified beliefs will apply as well to inferential justification. If the skeptic were to succeed in convincing the externalist that we are not noninferentially justified in believing propositions about the physical world, for example, the externalist presumably would argue that such beliefs are inferentially justified. The reliabilist, for example, would argue that if our beliefs about the external world result from input that includes beliefs about the internal and external conditions of perceiving, or even beliefs about the qualitative character of sensation, the relevant belief-dependent processes are conditionally reliable and therefore produce (inferentially) justified beliefs, *provided that the input beliefs are themselves justified*. The proviso is crucial, of course, and reminds us that to establish that first-level skepticism is false, the externalist who concedes that the justification is inferential in character must establish the existence of at least one unconditionally reliable process and at least one conditionally reliable process.

We noted in discussing the externalist's views about noninferentially justified belief that externalism has a potentially significant advantage in dealing with skepticism precisely because there are no restrictions on how large the class of noninferentially justified beliefs might be. As I indicated, there is no a priori reason for the externalist to deny even that we have noninferentially justified beliefs about the past and the future. Evolution might have taken care of us rather well when it comes to reaching true conclusions about the world, and evolution might have accomplished this end without burdening our brains with too many conditionally reliable belief-forming processes. Nozick's tracking relations can in principle hold between any fact and any belief, and the tracking relations *need* not involve any intermediate beliefs.

Just as the externalist's class of noninferentially justified beliefs can be very large in comparison to those recognized by traditional foundationalists, so the class of inferences recognized as legitimate by the externalist can be equally large. Consider again the reliabilist's position. There are no a priori restrictions on how many different kinds of conditionally reliable belief-dependent processes there might be. Valid deductive inference is presumably the paradigm of a conditionally reliable belief-dependent process. Classical enumerative induction may satisfy the requirements as well, provided that we find some suitably restricted

characterization of the inductive "process" that succeeds in denoting and that takes care of grue/green riddles of induction.⁸ I suspect most externalists will be reluctant to include perceptual beliefs among the beliefs produced by belief-dependent processes, but there is no reason why a reliabilist could not be a sense-datum theorist or an appearing theorist who holds that we do have at least dispositional beliefs about the qualitative character of sensation and who further holds that such beliefs are processed by conditionally reliable belief-dependent processes that churn out commonsense beliefs about the physical world. In short, take any kind of inference that people actually make and the reliabilist could hold that it involves a conditionally reliable belief-dependent process. All one needs to do is to formulate a description of the process that takes the beliefs one relies on as premises (the input) and produces the beliefs that constitute the conclusion (the output). The description will have to be such that we succeed in picking out a *kind* of process that does play the causal role described, but it will not need to involve any reference to the "hardware" of the brain. Indeed, we can try to *denote* the relevant process by directly referring only to the kind of premises and conclusion with which it is associated. Roughly, the idea is that we can try to denote a belief-dependent process *X*, for example, using the description "the process (whatever it is) that takes premises like these and churns out conclusions like this." Of course, such a description is probably too vague to do the trick. The locution "like these" can hardly be said to characterize precisely enough a class of premises. One would need to characterize the relevant points of similarity to have a well-defined class of premises which could then enter into the definite description denoting the process that takes them as input.

If we consider any argument someone actually makes, there will be indefinitely many classes of propositions to which the premises and the conclusion belong, and that will enable us to formulate any number of different descriptions of belief-forming processes. This is *not* a difficulty for the reliabilist, for as long as we have a locution that succeeds in denoting a process playing a causal role, we can use conditionals to define the conditions under which it is or is not conditionally reliable. The fact that a single inference might be subsumed under a number of different reliable belief-dependent processes is hardly a problem. If

the inference can be subsumed under the description of both a reliable and an unreliable process, the crucial question will be which process is causally determining the production of a belief. Thus, if someone trustworthy tells me today that it rained in New York, I can describe this as a case of processing testimony to reach a conclusion about the truth of what is testified to, or I can describe it as a case of taking a statement I hear involving the name "New York" and believing all of the noun clauses containing that name. The former, let us suppose, is a reliable belief-dependent process, whereas the latter is not. But you recall that in formulating descriptions of processes appealing to kinds of premises and conclusions, we are merely *hoping* to denote some process (presumably a complex brain process) that does take input and causally produce output beliefs. It does not follow, of course, that every definite description we formulate will succeed in denoting. In the hypothetical situation we are discussing, it may be that there is no programming in the brain that takes the "New York" input and processes it in the way described. If there is nothing denoted by the description playing the relevant causal role, then we do not need to worry about the fact that such a process, *if used*, would be unreliable.

To emphasize the point made earlier, according to externalism there are indefinitely many candidates for legitimate inferential processes. There are no a priori restrictions on how many conditionally reliable belief-dependent processes might be operating in normal human beings. There are no a priori restrictions on how many belief-dependent tracking relations might exist between beliefs and the facts that they track. Furthermore, just as in the case of noninferential justification, the question of which inferential processes generate justified beliefs for the externalist will be a purely contingent fact of a sort inaccessible to most philosophers *qua* philosophers. The existence of conditionally reliable processes, tracking relations, and the like is something that could be discovered only as a result of empirical investigation into causal relations. Philosophers are not trained to engage in this sort of empirical investigation.

Externalism, Normative Epistemology, and the Limits of Philosophy

Based on the observations above, I argue that if externalist metaepistemologies are correct, then

normative epistemology is an inappropriate subject matter for philosophy. Philosophers as they are presently trained have no special *philosophical* expertise enabling them to reach conclusions about which beliefs are or are not justified. Since the classic issues of skepticism fall under normative epistemology, it follows that if externalism were correct, philosophers should simply stop addressing the questions raised by the skeptic. The complex causal conditions that determine the presence or absence of justification for a belief are the subject matter of empirical investigations that would take the philosopher out of the easy chair and into the laboratory.

The realization that a good part of the history of epistemology becomes irrelevant to contemporary philosophy if we become metaepistemological externalists might cause a good many philosophers to reconsider externalism. I have always found the skeptical challenge to be fascinating and it has always seemed to me that I can address the relevant issues from my armchair (or my bed, depending on how lazy I happen to feel on a given day). If I had wanted to go mucking around in the brain trying to figure out the causal mechanisms that hook up various stimuli with belief, I would have gone into neurophysiology.

To rely on the philosopher's interest in skepticism and penchant for armchair philosophy as a rhetorical device to convert potential externalists, however, might be viewed as a new low in the art of philosophical persuasion. The mere fact that philosophers have been preoccupied with a certain sort of question does not mean that they were qualified to answer it. There are all kinds of perfectly respectable candidates for misguided philosophical investigations. Many philosophers, for example, have taken the question of whether every event has a cause to be a deep metaphysical issue in philosophy. As a good Humean, I would be the first to argue that it is a purely contingent question and if one wants to know the answer to it, one should not ask a philosopher.

Analogously, the fact that philosophers have been preoccupied with the skeptical challenge for literally thousands of years should not stop contemporary epistemologists from entertaining the thesis that the appropriate subject matter of epistemology ends with metaepistemology. After the metaepistemological analysis is complete, the externalist might argue, the only way to answer normative questions in epistemology is to engage in the kind of empirical investigation that contemporary

philosophers have not been trained *by philosophy* to do.

In reaching this conclusion I should be careful to admit that the philosophical externalist can, of course, embed normative epistemological conclusions in the consequents of conditional assertions. One can talk about what one would be justified in believing were certain conditions to obtain. But these conditionals are still part of metaepistemology. Indeed, such conditionals are merely a way of illustrating the consequences of metaepistemological positions as they apply to particular hypothetical situations. A Nozick, for example, can discuss what one would or would not know about the external world *if* a tracking analysis of knowledge were correct and *if* our beliefs about the physical world track the facts that would make them true. Nozick's analysis of knowledge also has the interesting feature that we can apparently determine a priori that we do *not* know certain things, for example, that we do not know that there is no evil demon deceiving us. But there will be no positive normative claim with respect to empirical knowledge that Nozick is particularly competent to make *qua philosopher*. As we shall see in a moment, externalism does not prevent a philosopher from reaching rational conclusions about what one is justified in believing. My conclusion is only that a philosopher's philosophical expertise is nothing that helps in reaching such conclusions. To illustrate this claim more clearly, let us turn to the question of whether externalist metaepistemologies suggest that one should be a skeptic about whether or not one has justified belief.

Second-Level Skepticism and the Fundamental Problem with Externalism

It is tempting to think that externalist analyses of justified or rational belief and knowledge simply remove one *level* the traditional problems of skepticism. When one reads the well-known externalists, one is surely inclined to wonder why they are so sanguine about their supposition that our commonsense beliefs are, for the most part, justified, if not knowledge. When Nozick, for example, stresses that interesting feature of his account allowing us to conclude consistently that we know that we see the table even though we do not know that there is no demon deceiving us, we must surely wonder *why* he is so confident that the subjunctives that on his view are sufficient for knowledge

are true. Perception, memory, and induction *may* be reliable processes in Goldman's sense, and thus given his metaepistemological position we *may* be justified in having the beliefs they produce, but, the skeptic can argue, we have no reason to believe that these processes *are* reliable, and thus, even if we accept reliabilism, we have no reason to conclude that the beliefs they produce are justified.

In the previous section I emphasized that if externalism is true then philosophers *qua philosophers* may not be particularly competent to answer normative questions in epistemology. I did *not* assert that if externalism is true we have no reason to believe that we have justified belief in commonsense truths about the world around us. According to externalist epistemologies, it is a purely contingent question as to what kinds of beliefs are justified. The existence of justified beliefs depends on nomological features of the world – facts about the reliability of belief-producing processes, the existence of tracking relations, causal connections between facts and beliefs, and the like. There are no a priori restrictions on what one might be justified in believing. But it *follows* from this that there are also *no* a priori restrictions on second-level knowledge or justified belief. It will also be a purely contingent question as to whether we have knowledge of knowledge or justified beliefs about justified beliefs. If we accept the externalist's metaepistemological views, it *may* be true that not only do we know what we think we know, but we also know that we know these things. Similarly, we may not only have all the justified beliefs we think we have, but we might also be justified in believing that we have these justified beliefs. The processes that yield beliefs about reliable processes may themselves be reliable. The beliefs about the truth of the subjunctives that Nozick uses to define first-level knowledge might themselves be embedded in true subjunctive conditionals that, *given the metaepistemological view*, are sufficient for second-level knowledge. My belief that my belief that *P* tracks the fact that *P* might track the fact that my belief that *P* tracks the fact that *P*. And there is no greater problem in principle when we move up levels. A reliable process might produce a belief that a reliable process produced the belief that my belief that *P* was produced by a reliable process. There might be a tracking relation tracking the tracking relation that tracks the fact that my belief that *P* tracks the fact that *P*. To be sure, the sentences describing the

conditions for higher levels of metajustification might look more like tongue-twisters than meta-epistemological analyses but, as ugly as they are, they are perfectly intelligible, and there is no *a priori* reason why the conditions required for higher-level justified belief and knowledge might not be satisfied.

It is also important to note that according to the externalist, in order to be justified in believing that I have a justified belief that *P*, I need not know anything about the *details* of the nomological connections sufficient for knowledge. Consider again reliabilism. In order to be justified in believing that my belief that *P* is produced by a reliable process, I do not need to know the physiological details of the brain states linking stimuli and belief. I would need to believe that there is *some* process producing the belief and I would need to believe that the process is reliable, but I would not need to know very much about what that process is. As I indicated earlier, one can denote the processes that produce beliefs using definite descriptions that refer directly only to the kinds of premises and conclusions that are linked by the process. Of course, the definite descriptions might fail to denote, and the beliefs in propositions expressed using such definite descriptions will either be false or meaningless (depending on what one does with the truth value of statements containing definite descriptions that fail to denote). But the descriptions might be successful, and in any event the belief that there is a reliable process taking stimuli *S* and resulting in belief *P* might itself be produced by a reliable process.

All this talk about what would in principle be possible given an externalist metaepistemology is fine, the skeptic might argue. But *how* exactly would one justify one's belief that, say, perception and memory are reliable processes? The rather startling and, I think, disconcerting answer is that *if* reliabilism is true, and *if* perception happens to be reliable, we could *perceive* various facts about our sense organs and the way in which they respond to the external world. Again, *if* reliabilism is true, and *if* memory is reliable, we could use memory, in part, to justify our belief that memory is reliable. You want a solution to the problem of induction? There is potentially no difficulty for the externalist. If reliabilism is true, and if inductive inference is a conditionally reliable belief-dependent process, then we can inductively justify the reliability of inductive inference. Our inductive justification for the reliability of inductive infer-

ence might itself be reliable, and if it is, that will give us second-level justification that our inductive conclusions are justified. A solution to the problem of induction will be important because with induction giving us inferentially justified conclusions, we can use inductive inference with the deliverances of perception and memory to justify our belief that those processes are reliable. I can remember, for example, that I remembered putting my keys on the desk and I can remember the keys being on the desk. If memory is an unconditionally reliable belief-independent process, then both my belief that I remembered putting the keys on the desk and my belief that I put the keys on the desk will be justified. I now have a premise that can be used as part of an inductive justification for memory being reliable. The more occasions on which I can remember memory being reliable, the stronger my inductive argument will be for the general reliability of memory.

The skeptic could not figure out how to get from sensations to the physical world. Assume that perception is itself a belief-independent, unconditionally reliable process. Assume also that whatever perception involves, its specification involves reference to sensation, and assume further that we have "introspective" access to sensation. Introspective access might itself be another belief-independent, unconditionally reliable process. Given these suppositions, if reliabilism is true, then introspection can give us justified beliefs that we are perceiving, and perception can give us justified beliefs that physical objects are present. The two reliable processes together can furnish a premise that, when combined with others generated in a similar fashion, gives us inductive justification for believing that perception is reliable. So if both introspection and perception happen to be reliable, there seems to be no great obstacle to obtaining justified belief that they are reliable. Second-level justified belief is not much more difficult to get than first-level justified belief.

How successful *inductive* reasoning will be in answering second-level skeptical questions depends very much on how the externalist resolves some of the controversies discussed elsewhere, specifically on how narrowly the relevant belief-forming processes are characterized. I have pointed out that as long as reliability is not defined in terms of actual frequencies, there is no *conceptual* difficulty in a reliabilist positing the existence of very narrowly defined, reliable belief-forming

processes that have only a few, or even no, instances. Although there is no conceptual difficulty in supposing that there are such processes, it obviously creates problems for any *inductive* justification for believing that they exist and are reliable. As should be clear by now, however, the unavailability of inductive justification in no way implies that there is not some *other* reliable belief-forming process that will still yield second-level knowledge or justified belief.

This reminds us, of course, of Quine's injunction to naturalize epistemology.⁹ Quine suggested that we give ourselves full access to the deliverances of science when it comes to understanding how we have knowledge of the world around us. Contemporary externalists have simply given us more detailed metaepistemological views which allow us to rationalize following the injunction to naturalize epistemology. If the mere reliability of a process, for example, is sufficient to give us justified belief, then *if* that process is reliable we can use it to get justified belief wherever and whenever we like.

All of this will, of course, drive the skeptic crazy. You cannot *use* perception to justify the reliability of perception! You cannot *use* memory to justify the reliability of memory! You cannot *use* induction to justify the reliability of induction! Such attempts to respond to the skeptic's concerns involve blatant, indeed pathetic, circularity. Frankly, this does seem right to *me* and I hope it seems right to *you*, but *if* it does, then I suggest that you have a powerful reason to conclude that externalism is false. I suggest that, ironically, the very ease with which externalists can deal with the skeptical challenge at the next level betrays the ultimate implausibility of externalism as an attempt to explicate concepts that are of *philosophical* interest. If a philosopher starts wondering about the reliability of astrological inference, the philosopher will not allow the astrologer to read in the stars the reliability of astrology. Even if astrological inferences happen to be reliable, the astrologer is missing the point of a *philosophical* inquiry into the justifiability of astrological inference if the inquiry is answered using the techniques of astrology. The problem is perhaps most acute if one thinks about first-person philosophical reflection about justification. If I really am interested in knowing whether astrological inference is legitimate, if I have the kind of philosophical curiosity that leads me to raise this question in the first place, I will not for a moment suppose that further

use of astrology might help me find the answer to my question. Similarly, if as a philosopher I start wondering whether perceptual beliefs are accurate reflections of the way the world really is, I would not dream of using perception to resolve my doubt. Even if there is some sense in which the reliable process of perception might yield justified beliefs about the reliability of perception, the use of perception could never satisfy a *philosophical curiosity* about the legitimacy of perceptual beliefs. When the philosopher wants an answer to the question of whether memory gives us justified beliefs about the past, that answer cannot possibly be provided by memory.

Again, if one raises skeptical concerns understanding fundamental epistemic concepts as the externalist does, then there should be no objection to perceptual justifications of perception, inductive justifications of induction, and reliance on memory to justify the use of memory. If one is understanding epistemic concepts as the reliabilist suggests, for example, then one can have no objection in principle to the use of a process to justify its use. After all, the whole point of inferential externalism is to deny the necessity of having access to the probabilistic relationship between premises and conclusion in order to have an inferentially justified belief. The mere reliability of the process is sufficient to generate justified belief in the conclusion of an argument. There is no conceptual basis for the reliabilist to get cold feet when epistemological questions are raised the next level up. Either reliability alone is sufficient or it is not. If it is, then it is sufficient whether one is talking about justification for believing *P* or justification for believing that one has a justified belief that *P*.

It is both interesting and illuminating that even many access externalists seem to worry about the possibility of second-level justification in ways that they do not worry about the possibility of first-level justification. Alston explicitly rejects the idea that one needs access to the adequacy of one's grounds for believing *P* in order to be justified in believing *P*. But in *The Reliability of Sense Perception*, he also seems to reject the idea that one can use a "track record" argument (an inductive argument of the sort I sketched above) to *justify* one's belief that perception and memory are reliable. Such arguments will inevitably presuppose the adequacy of the very grounds whose adequacy is at issue. In doing so the argument will be viciously circular.

But what exactly is Alston's complaint? To justify my belief that perception or memory is reliable, I need only find a good argument whose premises I justifiably accept and whose premises support the conclusion that these ways of forming beliefs are reliable. But if perception and memory are reliable *and there is no requirement of access to adequacy of grounds* in order for a belief to be justified, what is the problem? Why is it harder to justify my belief that my perceptual beliefs are justified than it is to have justified beliefs based on perception?

Much of the time Alston seems to admit everything I have just said, but the circular nature of the available arguments still clearly bothers him:

if sense perception is reliable [Alston's emphasis], a track record argument will suffice to show that it is. Epistemic circularity does not in and of itself disqualify the argument. But even granting that point, the argument will not do its job unless we *are* justified in accepting its premises; and this is the case only if sense perception is in fact reliable.

... But when we ask whether one or another source of belief is reliable, we are interested in *discriminating* those that can reasonably be trusted from those that cannot. Hence merely showing that if a given source is reliable it can be shown by its record to be reliable, does nothing to indicate that the source belongs with the sheep rather than with the goats. (p. 17)

But again, *as an externalist*, what does Alston want? He obviously thinks that in some sense all we could ever really conclude is that we *might* have justification for thinking that we have justified beliefs based on perception. And the contextual implication of this claim is that we also *might not*. But what "might" is this? Clearly, it is intended to refer to *epistemic* possibility. Let us say that *P* is epistemically possible for *S* when *P* is consistent with everything that *S* knows. Is it epistemically possible for us that perception is unreliable? Not if perception is reliable, because we will have inductive knowledge that it is not unreliable. But you are still just asserting a conditional, Alston will complain. For all we know, it is possible that perception is unreliable. But this claim about epistemic possibility is precisely the claim that Alston, as an externalist, has no business making. Can we *discriminate* (his word) between reliable and unreliable fundamental sources of belief? As an

externalist he has no reason to deny that we can and do discriminate between reliable and unreliable processes (using, of course, reliable processes). Alston clearly wants to assert (and assert justifiably) a conclusion about epistemic possibility. But the concept of epistemic possibility he wants to apply at the second level is not one that can be understood within the framework of the externalism he embraces.

I agree, of course, with Alston's conclusion that one cannot use perception to justify one's belief that perception is reliable and memory to justify one's belief that memory is reliable. But that is only because the externalist is wrong in characterizing the concept of justification that even externalists are often interested in when they move up levels and start worrying about whether they can justify their belief that their beliefs are justified. The epistemic concept of discrimination that Alston invokes in the passage I quoted is precisely the concept that is at odds with his own attempt to defend an externalist understanding of epistemic concepts.

The fundamental objection to externalism can be easily summarized. If we understand epistemic concepts as the externalists suggest we do, then there would be no objection in principle to using perception to justify reliance on perception, memory to justify reliance on memory, and induction to justify reliance on induction. But there is no philosophically interesting concept of justification or knowledge that would allow us to use a kind of reasoning to justify the legitimacy of using that reasoning. Therefore, the externalist has failed to analyze a philosophically interesting concept of justification or knowledge.

The objection is by no means decisive. Obviously, many externalists will bite the bullet and happily embrace Quine's recommendation to naturalize epistemology. If the argument convinces anyone, it will be those who were initially inclined to suppose that externalism will inevitably encounter skepticism at the next level up. Maybe we have knowledge or justified belief as the externalist understands these concepts, some would argue, but we would never be in a position to know that we have knowledge or justified belief if the externalist is right. The only reason I can see for granting the first possibility but denying the second is that one is implicitly abandoning an externalist analysis of epistemic concepts as one moves to questions about knowledge or justification at the next level. But if when one gets philosophically "serious" one

abandons the externalist's understanding of epistemic concepts, then, for philosophical purposes, one should not concede the externalist's understanding of epistemic concepts at the first level. Once you concede that according to the externalist we might have knowledge or justified belief about the past and the external world, you have also implicitly conceded that we might have knowledge that we have such knowledge, justified belief that we have such justified belief. And we might also have knowledge that we have knowledge that we

have knowledge, and have justified beliefs that we have justified beliefs that we have justified beliefs. It seems to many of us that the externalist is simply missing the point of the philosophical inquiry when externalist analyses of epistemic concepts continue to be presupposed as the skeptical challenge is repeated at the metalevels. But the only explanation for this is that the externalist analysis of epistemic concepts never was adequate to a philosophical understanding of epistemic concepts.

Notes

- 1 For a detailed defense of this last view, see Fumerton, *Metaphysical and Epistemological Problems of Perception* (Lincoln, NE: University of Nebraska Press, 1985).
- 2 I stress "qua philosopher" for there is a real danger that I will be misunderstood on this point. Later in this chapter I argue that externalism is perfectly compatible with philosophers (and anyone else) having justified beliefs about whether or not they have justified beliefs. It will not, however, be their philosophical competence that yields such justification.
- 3 As I shall argue shortly, this claim might be misleading. In one sense the detailed character of belief-forming processes would be best discovered by neurophysiologists. But there is another sense in which anyone can form beliefs about such processes, even without any detailed knowledge of how the brain works.
- 4 See Wilfrid Sellars, *Science Perception and Reality* (London: Routledge and Kegan Paul, 1963), pp. 146-7.
- 5 Roderick M. Chisholm, *Perceiving* (Ithaca, NY: Cornell University Press, 1957), p. 49.
- 6 Again the terminology and the distinction is introduced by Chisholm in *ibid.*, pp. 50-3.
- 7 It should go without saying that there may be causally necessary conditions for the existence of such non-inferential justification having to do with the capacity to form other beliefs. The dependency that concerns us, however, is logical. Justification is noninferential when no other belief is a *constituent* of the justification.
- 8 The allusion is, of course, to the problem discussed in Nelson Goodman, *Fact, Fiction and Forecast* (Indianapolis: Bobbs-Merrill, 1955), ch. 3.
- 9 W. V. Quine, *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), ch. 3.

Knowledge and the Internal

John McDowell

1. I am going to work with an idea from Sellars, that knowledge – at least as enjoyed by rational animals – is a certain sort of standing in the space of reasons.¹ My concern is a familiar philosophical dialectic, which I shall approach in terms of what happens to the Sellarsian idea when the image of standings in the space of reasons undergoes a certain deformation. That it is a deformation is something we can learn from how unsatisfactory the familiar dialectic is.

2. The deformation is an interiorization of the space of reasons, a withdrawal of it from the external world. This happens when we suppose that we ought to be able to achieve flawless standings in the space of reasons by our own unaided resources, without needing the world to do us any favors.

Consider the Argument from Illusion. Seeing, or perhaps having seen, that things are thus and so would be an epistemically satisfactory standing in the space of reasons. But when I see that things are thus and so, I take it that things are thus and so on the basis of having it look to me as if things are thus and so. And it can look to me as if things are thus and so when they are not: appearances do not give me the resources to ensure that I take things to be thus and so on the basis of appearances only when things are indeed thus and so. If things are indeed thus and so when they seem to be, the world is doing me a favor. So if I want to restrict myself to standings in the space of reasons whose flawlessness I can ensure without external help, I must go no further than taking it that it looks to me as if things are thus and so.

Originally published in *Philosophy and Phenomenological Research* 55 (1995), pp. 877–93.

One might hope that this inward retreat is only temporary. Take a particular case in which it looks to me as if things are a certain way. If things are indeed that way, that is – so far – a favor the world is doing me. The hope is that I might start from the fact that things look that way to me; add in anything else that the ground rules allow me to avail myself of, if it helps, and move from there, by my own unaided resources, without needing the world to do me any favors, to a satisfactory standing in the space of reasons with respect to the fact that the world is arranged the way it looks. And now that would no longer be a favor the world is doing me, a kindness I must simply hope for. Now I would have a derivatively satisfactory standing in the space of reasons, with respect to the fact that things are as they look, which I achieved by myself without needing to be indebted to the world.

Anyone who knows the dreary history of epistemology knows that this hope is rather faint. That will matter in due course, but it does not matter for what I am doing now, which is simply reminding you, in perhaps slightly unfamiliar terms, of a familiar epistemologists' syndrome.

Perception could not yield us standings in the space of reasons at all without some indebtedness to the world. The position I am describing does not involve the fantasy that pure unaided reason could give us knowledge of the external world, without our needing the world to oblige us by affording appearances. The thought is that we need no outside help in avoiding being led astray by whatever appearances the world is kind enough to afford us.

The Argument from Illusion is of course familiar in the epistemology of perception. But an

argument with the same structure is tempting whenever it can seem right to say that a favor from the world is needed if there is to be application for a locution of the relevant kind: a locution that belongs with “see that . . .” in that it is epistemic and thereby factive. This will be so whenever appearances can be misleading, in such a way that the potential deception cannot be blamed on defects in how one has conducted oneself in the space of reasons. What is at fault must then be the unkindness of the world. And when an appearance is not misleading, that is, correspondingly, a favor from the world. Whenever we have that structure, it will seem that the epistemic position signalled by the original locution can be at best derivative; the true starting-point in the space of reasons must be something common to the favorable and the potentially misleading cases (like having it look to one as if things are thus and so).²

Factive locutions that are vulnerable to this treatment include “remember that . . .” (with “seem to remember that . . .” as the upshot of the retreat), and “learn from so-and-so that . . .” (with “hear so-and-so say that . . .” as the upshot of the retreat).³ A negative instance may help to make the point clear: consider “prove that . . .” Suppose one is subject to a misleading appearance that one has a proof of something. In that case, surely, one must have misconducted oneself in the space of reasons; it cannot be that the world is the only thing one can blame for what has gone wrong.⁴

3. I spoke of hoping that the inward retreat is only temporary, and I suggested that the hope is faint. I think this is true in all the applications, but I shall stick to perception to bring out what this implies.

One need not restrict oneself to the particular perceptual appearance whose credentials are in question. I allowed that one could add in anything else that might help, if it is available according to the ground rules. Here we might think of surrounding appearances and background knowledge. (It will emerge that it is open to doubt whether the ground rules make any background knowledge available, but we can let that pass at this stage.)

Clearly one is not stuck with simply believing, come what may, that things are as they appear. One can refine one’s policies or habits of basing beliefs on appearance, taking more and more circumstances into account, with a view to improving the proportion of truths to falsehoods in their output. And it is not just that one *can* engage in this refining procedure. Surely reason positively

requires one to do so. If it turns out to be an effect of interiorizing the space of reasons that we become unable to make sense of this critical function of reason, we ought to conclude that the very idea of the space of reasons has become unrecognizable. I think that is what turns out: I want to bring that out by giving the idea of something that is both interiorized and still recognizably the space of reasons a run for its money.

So we are to try to reconstruct the epistemic satisfactoriness implicit in the idea of seeing that things are thus and so, using the following materials: first, the fact that it looks to a subject as if things are that way; second, whatever further circumstances are relevant (this depends on the third item); third, the fact that the policy or habit of accepting appearances in such circumstances is endorsed by reason, in its critical function, as reliable. And now the trouble is this: unless reason can come up with policies or habits that will *never* lead us astray, there is not enough here to add up to what we were trying to reconstruct. Seeing that things are thus and so is a position that one cannot be in if things are not thus and so. Given that one is in that position, it follows that things are thus and so. And if reason cannot find policies or habits that are utterly risk-free, the reconstructing materials cannot duplicate that. However careful one is in basing belief on appearances, if one’s method falls short of total freedom from risk of error, the appearance plus the appropriate circumstances for activating the method cannot ensure that things are as one takes them to be.

There are various possible responses to this point. The one I recommend is that we should jettison the whole approach to knowledge that structures epistemology around the Argument from Illusion. I shall mention three others.

Obviously one response is skepticism. In my Sellarsian framework, I can put the skeptical response like this. An epistemically satisfactory position would have to be a standing in the space of reasons – Sellars is right about that. But the argument I have just sketched shows that we cannot reconstruct a standing in the space of reasons, suitable to amount to knowledge, with respect to the fact that things are as they perceptually appear. So it must be a mistake to think we can achieve knowledge through perception. This thought clearly generalizes, in a way that matches the generalization of the Argument from Illusion.

A second response would be to claim that there must be policies or habits of basing belief on

appearance that *are* utterly risk-free. It is obvious how this response might be attractive, in the context of the threat of skepticism; but I do not think it has any plausibility in its own right. It would express a rather touching *a priori* faith in the power of human reason to devise fully effective protections against the deceptive capacities of appearance. No doubt it would suit our vanity, or at least help us feel safe, if we could suppose our reason had such power, but obviously that is no ground to believe it is so; in fact it is a ground to be suspicious of the idea. (I shall return to a point of this sort later.)⁵

A third response is to keep the Sellarsian idea I began with, in its interiorized form, but only as one element in a composite conception of knowledge; we are to add an external element in order to cope with the problem I am considering. The upshot is a position that looks like this. At least for rational animals, a satisfactory standing in the space of reasons is a necessary condition for knowledge. But since the positions that one can reach by blameless moves in the space of reasons are not factive, in the way that epistemically satisfactory positions are, a satisfactory standing in the space of reasons cannot be what knowledge is.⁶ Rather, knowledge is a status that one possesses by virtue of an appropriate standing in the space of reasons when – this is an extra condition, not ensured by one's standing in the space of reasons – the world does one the favor of being so arranged that what one takes to be so is so.

It calls for comment that the external addition I am envisaging is the familiar truth requirement for knowledge, that what one takes to be so is indeed so. This figures, in the position I am considering, as a necessary extra condition for knowledge, over and above the best one can have in the way of reliability in a policy or habit of basing belief on appearance. Such reliability figures in the hybrid position I am considering as part of its internal apparatus: as something to be taken into account when one determines, within the space of reasons conceived in the interiorized way, whether a standing in that space is acceptable.

This internal placement of reliability may seem surprising, in view of the fact that reliability is part of the stock in trade of full-blown externalist approaches to knowledge. But the point of full-blown externalist approaches is to reject the Sellarsian idea that I began with, not to incorporate it as part of an account of knowledge, as in the hybrid approach I am considering. According to

a full-blown externalist approach, knowledge has nothing to do with positions in the space of reasons: knowledge is a state of the knower linked to the state of affairs known in such a way that the knower's being in that state is a reliable indicator that the state of affairs obtains. In the purest form of this approach, it is at most a matter of superficial idiom that we do not attribute knowledge to properly functioning thermometers.⁷ Now from the fact that the concept of reliability plays this external role in an approach that simply rejects questions about the knower's position in the space of reasons, it clearly does not follow that when we move to an approach that does not reject such questions (although, being hybrid, it insists that they do not exhaust the issues that need to be addressed), the concept must still be conceived as operating outside the space of reasons. And the point I made earlier stands: if a purported picture of the space of reasons makes no room for the critical function of reason in raising questions about the reliability of this or that policy or habit of belief-formation, the picture cannot be what it purports to be. So it would be a mistake to suppose that reliability must be external in a hybrid approach, just because it figures in full-blown externalist approaches. Reliability must operate in the internal reaches of a hybrid approach, on pain of the internal element's becoming unrecognizable as what it was supposed to be.⁸ The problem with the resources that are available in an interiorized conception of the space of reasons is that, even including the best that can be had in the way of reliability, they cannot duplicate the factiveness of epistemically satisfactory positions; so it is precisely the truth requirement that these considerations motivate conceiving as an external condition that needs to be added to internal requirements for knowledge.⁹

4. This hybrid conception of knowledge has an evident instability, in the way it separates truth, which figures as an external element, from reliability in policies or habits of belief-formation, which figures as an internal element. The truth requirement has to be an added external element, because the interiorization of the space of reasons means that there cannot be standings in that space that simply consist in a cognitive purchase on an objective fact, for instance something that one perceives to be so, or remembers to have been so. But if there cannot be such standings in the space of reasons, how can reason have the resources it

would need in order to evaluate the reliability of belief-forming policies or habits? If we press this question, the idea that something can be both interiorized in the way I am considering and recognizably a conception of the space of reasons starts to unravel, as I have already hinted that it would.

I shall return to that point; meanwhile I want to urge another problem about the hybrid conception of knowledge. In the hybrid conception, a satisfactory standing in the space of reasons is only part of what knowledge is; truth is an extra requirement. So two subjects can be alike in respect of the satisfactoriness of their standing in the space of reasons, although only one of them is a knower, because only in her case is what she takes to be so actually so. But if its being so is external to her operations in the space of reasons, how can it not be outside the reach of her rational powers? And if it is outside the reach of her rational powers, how can its being so be the crucial element in an intelligible conception of her knowing that it is so – what makes the relevant difference between her and the other subject? Its being so is conceived as external to the only thing that is supposed to be epistemologically significant about the knower herself, her satisfactory standing in the space of reasons. That standing is not itself a cognitive purchase on its being so; it cannot be that if the space of reasons is interiorized. But then how can the unconnected obtaining of the fact have any intelligible bearing on an epistemic position that the person's standing in the space of reasons is supposed to help constitute? How can it coalesce with that standing to yield a composite story that somehow adds up to the person's being a knower?

One way to appreciate what I am driving at here is to consider the familiar point that true belief need not amount to knowledge. Why not? A good simple answer is that mere truth in a belief leaves it open that the believer has hold of the truth by accident, and knowledge excludes that. Now in the hybrid conception of knowledge, it is admittedly not a complete accident, relative to someone's standing in the space of reasons, if things are as she takes them to be: the position of her belief in the space of reasons makes it likely to be true. But the reason why the extra stipulation that the belief is true – what is distinctive of the hybrid approach – is needed is that likelihood of truth is the best that the space of reasons yields, on the interiorized conception of it: the closest we can come to factiveness. The extra that we need for knowledge – the fact that the case in question is not one of those

in which a largely reliable habit or policy of belief-formation leads the subject astray – is, relative to the knower's moves in the space of reasons, a stroke of good fortune, a favor that the world does her. So if we try to picture epistemic status as constituted in the way the hybrid conception has it, we are vulnerable to a version of the familiar point that distinguishes knowledge from mere true belief.

I think the moral of this is that if we cannot see our way to accepting the Sellarsian idea in full, we should reject it, as in full-blown externalist accounts. It is not a good idea to suppose that a satisfactory standing in the space of reasons might be part but not the whole of what knowledge is.

5. A hybrid conception of knowledge is often taken to be obvious.¹⁰ But in the light of what I have just argued, I think this depends on not thinking directly about the conception's epistemological credentials. What makes the hybrid conception seem obvious is that, leaving aside the full-blown externalism according to which standings in the space of reasons are irrelevant to knowledge, this view of knowledge seems to be the only alternative to skepticism. But this is one of those set-ups that are familiar in philosophy, in which a supposedly exhaustive choice confers a spurious plausibility on a philosophical position. The apparent plausibility is not intrinsic to the position, but reflects an assumed framework; when one looks at the position on its own, the plausibility crumbles away, or so I have tried to suggest. In such a situation, the thing to do is to query the assumption that seems to force the choice. And in this case, the culprit is the interiorized conception of the space of reasons.

I have described that conception in a way that equips it with an intelligible motivation. The aim is to picture reason as having a proper province in which it can be immune to the effects of luck; not in the sense of sheer chance, but in the sense of factors that reason cannot control, or control for. The idea is that reason can ensure that we have only acceptable standings in the space of reasons, without being indebted to the world for favors received; if we exercise reason properly, we cannot arrive at defective standings in the space of reasons, in a way that could only be explained in terms of the world's unkindness.¹¹ The upshot of this interiorization is that knowledge of the external world cannot be completely constituted by standings in the space of reasons. The hybrid view concedes that such knowledge is partly a

matter of luck in the relevant sense, something outside the control of reason; the hope is that this admission of luck is tolerable, because it comes only after we have credited reason with full control over whether one's standings in the space of reasons are satisfactory.¹²

It seems clear where our suspicions should attach themselves. Although the motivation I have suggested for interiorizing the space of reasons is intelligible, it is surely something we ought to find suspect. The hybrid view's concession to luck, tagged on to a picture of reason as self-sufficient within its own proper province, comes too late; the very idea of reason as having a sphere of operation within which it is capable of ensuring, without being beholden to the world, that one's postures are all right – like the obvious analogues of this idea in thought about practical reason – has the look of a fantasy, something we spin to console ourselves for the palpable limits on our powers.¹³

To avoid fantasy, we would need to see our way to accepting that we cannot eliminate what the interiorized conception of reason conceives as a quite alien factor, the kindness of the world, as a contributor to our coming to occupy epistemically satisfactory positions in the space of reasons. This points to a different conception of factive positions like seeing that things are a certain way. When someone enjoys such a position, that involves, if you like, a stroke of good fortune, a kindness from the world; even so, the position is, in its own right, a satisfactory standing in the space of reasons, not a composite in which such a standing is combined with a condition external to the space of reasons.¹⁴ Whether we like it or not, we have to rely on favors from the world: not just that it presents us with appearances, which the fantasy view can already accept as a favor that the world does us, but that on occasion it actually is the way it appears to be. But that the world does someone the necessary favor, on a given occasion, of being the way it appears to be is not extra to the person's standing in the space of reasons. Her coming to have an epistemically satisfactory standing in the space of reasons is not what the interiorized conception would require for it to count as her own unaided achievement; but then once she has achieved such a standing, she needs no extra help from the world to count as knowing.¹⁵ If we rescue the idea of the space of reasons from the distortions of fantasy, we can say that the particular facts that the world does us the favor of vouchsafing to us, in the various relevant

modes of cognition, actually shape the space of reasons as we find it. The effect is a sort of coalescence between the idea of the space of reasons as we find it and the idea of the world as we encounter it.¹⁶

Of course we are fallible in our judgments as to the shape of the space of reasons as we find it, or – what comes to the same thing – as to the shape of the world as we find it. That is to say that we are vulnerable to the world's playing us false; and when the world does not play us false we are indebted to it.¹⁷ But that is something we must simply learn to live with, rather than recoiling into the fantasy of a sphere in which our control is total.¹⁸

6. The space of reasons is the space within which thought moves, and its topography is that of the rational interconnections between conceptual contents; we might equally speak of the space of concepts.¹⁹ So we can see the interiorization of the space of reasons as a form of a familiar tendency in philosophy: the tendency to picture the objective world as set over against a "conceptual scheme" that has withdrawn into a kind of self-sufficiency. The fantasy of a sphere within which reason is in full autonomous control is one element in the complex aetiology of this dualism.²⁰ The dualism yields a picture in which the realm of matter, which is, in so far as it impinges on us, the Given, confronts the realm of forms, which is the realm of thought, the realm in which subjectivity has its being. It is of course a second Sellarsian idea that this picture is hopeless; it is the source of the basic misconception of modern philosophy, that the task of philosophy is to bridge an ontological and epistemological gulf across which the subjective and the objective are supposed to face one another.²¹

This full-fledged dualism of subjective and objective – or inner and outer – is a good context in which to think about something I promised to come back to: the instability of an epistemology in which truth is external and reliability is internal.

When the dualism becomes full-fledged, it defeats itself. If we conceive what we want to think of as the space of concepts, the realm of thought, in a way that alienates it so radically from the merely material that we seem to be faced with those familiar modern problems of reconciling the subjective with the objective, we undermine our right to think of it as the realm of thought at all. When we set it off so radically from

the objective world, we lose our right to think of moves within the space we are picturing as content-involving. So we stop being able to picture it as the space of concepts. Everything goes dark in the interior as we picture it.²²

Now in the epistemological syndrome that I have been discussing, the aim is to set off the inner from the outer, but in a way that stops short of that disastrous extinguishing of content. The idea is that the outer injects content into the inner: the world affords us appearances, and we thereby have dealings with content (it seems to us that things are, or were, thus and so). Appearances are starting-points from which we can move about in the interior space, the space of reasons, drawing inferences from them in ways that reason can endorse, for instance on the ground that a particular inference exemplifies a mode of arriving at beliefs that is reliable. But the instability that I pointed to, the separation of truth as external from reliability as internal, reveals that this attempt to stop short of disaster is hopeless. If moves in the space of reasons are not allowed to start from facts, riskily accepted as such on the basis of such direct modes of cognitive contact with them as perception and memory, then it becomes unintelligible how our picture can be a picture of a space whose positions are connected by relations that reason can exploit, such as that one of them is a reliable ground for moving by inference to another. If the space of reasons as we find it is withdrawn from the objective world as it makes itself manifest to us, then it becomes unintelligible how it can contain appearances, content-involving as they must be, either. We are here in the vicinity of a third Sellarsian idea, that reality is prior, in the order of understanding, to appearance; I am drawing the moral that it makes no sense to suppose that a space sufficiently interiorized to be insulated from specific manifest fact might nevertheless contain appearances.²³

7. The considerations I have offered suggest a way to respond to skepticism about, for instance, perceptual knowledge; the thing to do is not to answer the skeptic's challenges, but to diagnose their seeming urgency as deriving from a misguided interiorization of reason. But at least one familiar form of skepticism is not obviously within reach of this move. At first appearance, at any rate, skepticism about induction does not turn on an interiorization of the space of reasons. In connection with inductive knowledge, we seem not to

need an Argument from Illusion to achieve the effect that the Argument from Illusion achieves in the cases where skeptics do appeal to it: the effect of focusing our attention on a basis – a starting-point in the space of reasons – that falls short of the facts supposedly known.

Without trying a full treatment, I shall mention a fourth Sellarsian idea whose effect is to bring inductive skepticism into the same framework. Consider a characteristic Humean formulation of the predicament that is supposed to invite inductive skepticism:

It may, therefore, be a subject worthy of curiosity, to enquire what is the nature of that evidence, which assures us of any... matter of fact, beyond the present testimony of our senses.²⁴

Taking it seriously that what is in question is testimony of our senses, we must think in terms of something content-involving – something in which, say, colors figure as apparent properties of objects. A mere wash of chromatic sensation, not referred to a supposedly perceived environment, could not count as testimony of our senses. Now my fourth Sellarsian idea can be put like this: there cannot be a predicament in which one is receiving testimony from one's senses but has not yet taken any inductive steps. To stay with the experience of color, whose simplicity presumably makes it maximally favorable to the contrary view: color experience's being testimony of the senses depends on the subject's already knowing a great deal about, for instance, the effect of different sorts of illumination on color appearances; and a subject could not know that without knowing a great deal, outside the immediate deliverances of the senses, about the objective world and our cognitive access to it.

This makes for an easy extension to inductive skepticism of the epistemological move I have been recommending. The key thought so far has been that if we refuse to make sense of the idea of direct openness to the manifest world, we undermine the idea of being in the space of reasons at all, and hence the idea of being in a position to have things appear to one a certain way. There is no making sense of perceptual appearances – the testimony of one's senses – without making sense of the possibility that the objective world can be immediately present to the senses. Now Sellars's point about color experience is a specific case of this point:

there is no making sense of that possibility unless one's conceptual space already embraces a world with more to it than is immediately present to the senses. Nothing could be immediately present to one's senses unless one already had knowledge that goes beyond what is immediately present to the senses. So the supposed predicament of the inductive skeptic is a fiction. And the mistake is really the same as the one I have already discussed: that we can make the inward withdrawal that the Argument from Illusion is supposed to compel, but stop short of extinguishing content. It is superficial to object that an Argument from Illusion does not typically figure in recommendations of inductive skepticism; Hume's formulation can seem to describe a predicament only if one does not think through the idea that its subject already has the testimony of the senses, and this means that skepticism about induction can seem gripping only in combination with a straightforwardly interiorizing epistemology for perception.

8. There may be a temptation to object that the interiorization I have been discussing cannot be a fantasy, as I have been suggesting it is; it is simply a version of a perfectly intuitive thought, a piece of common sense, to the effect that the mental is internal. If this intuitive thought is taken anything like literally, it can seem to compel the conclusion that minds make contact with the external world at an interface, and then cognitive states, with their factive nature, surely cannot but be composites of interior and exterior circumstances.²⁵ And a literal construal can be very tempting; after all, it is unquestionable that human beings do literally have insides, and that they are partly occupied by complex mechanisms about whose operations we can in principle, and to some extent in practice, do natural science, in such a way as to account in some sense for behavior – the very thing that appealing to the mental was supposed to do.

I cannot deal with this properly now; but I shall end by mentioning two reasons for doubt about this line of objection.

First, there is a familiar and impressive tradition of reflection about common-sense psychology, according to which the point of its concepts lies in their providing a kind of understanding of per-

sons and their doings that is radically unlike the understanding that the natural sciences can yield. This tradition's insights are never taken sufficiently seriously by proponents of the idea that "folk psychology" is a proto-theory of the operations of those internal mechanisms, to be refined and perhaps wholly superseded as we learn more about what goes on inside our heads. For instance, natural-scientific investigation of how what is literally internal controls behavior would seek theories whose power to explain would be proportional to their power to predict. But folk-psychological concepts can express a kind of understanding of a person that seems to have little or no relation to predictive power. And we do not find it any the worse for that, at least until methodologically inclined people try to put us on the defensive. If the understanding that common-sense psychology yields is *sui generis*, there is no reason to regard it as a primitive version of the understanding promised by a theory of inner mechanisms. The two sorts of understanding need not compete for room to occupy.

Second, about the intuitive idea that the mental is internal. I suppose this idea makes it natural, when we learn about advances in the scientific understanding of how our behavior is controlled by literally internal mechanisms, to suppose that that is what we must have had some dim conception of all along. But I think this is a confusion. At its most abstract, the content of the pre-theoretical notion of the mental as inner lies in such facts as that at least some mental states and occurrences, unlike perceptible states of affairs, are "internal accusatives" to the consciousness of their subjects.²⁶ But the character of the notion comes out more concretely in the idea that one can sometimes see what someone's mental state is by (as we say) looking into her eyes. And this idea carries its nature on its face: it is a picture, a piece of imagery.²⁷ (This is not something to be embarrassed about.) It has nothing to do with the idiotic thought that one can look through the eyes into the interior of a person's skull. There is no comfort to be derived here, by way of a literal construal of a piece of common sense, for the withdrawal of the mental from direct engagement with the world that is expressed by the differently figurative interiorization that has been my main target.

Notes

- 1 "In characterizing an episode or a state as that [better: one] of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says" (Wilfrid Sellars, *Science, Perception and Reality* (Routledge and Kegan Paul, London, 1963), p. 169). I put in the parenthetical qualification so as to allow that a concept of knowledge might be applied to non-rational animals too; but nothing in this paper will depend on that.
- 2 I am deliberately leaving the idea of blameworthiness, in one's moves in the space of reasons, unspecific. If someone arrives at a false belief from which she would have been deterred by some investigation she chose not to engage in because of its high cost and low probability of overturning the other evidence, is she blameworthy? Different answers are possible. But no reasonable interpretation of the idea of doxastic obligations could make falsehood in an empirical belief show, all by itself, that an obligation has not been met. That is the central insight (a genuine insight, even though it is typically mishandled) of the familiar genre of philosophy according to which empirical knowledge is problematic. I want to focus on this gap, nearly universally acknowledged, between doxastic blamelessness in a sense that connects with doxastic obligation, on the one hand, and empirical knowledge, on the other, without being distracted by details about how doxastic blamelessness should be understood. (The epistemological outlook I shall recommend should make such questions seem less urgent.)
- 3 In the case of testimony, one is literally done a favor by an informant. But there will have to be, figuratively, a favor from the world at some point in the epistemic ancestry of a piece of knowledge by testimony. (At least outside the area of, for instance, being told of something that has been proved; see the text below.)
- 4 This is essentially the feature of proof (or computation) that Crispin Wright aims to generalize, in his account of what it is to have verified a statement ("Strict Finitism," *Synthese* LI (1982), pp. 203–82, at pp. 210–18). Wright strangely combines an understanding of this feature of proof (or computation) with applying the Argument from Illusion even here. He writes (p. 210): "If arithmetical computation is to be a paradigm of verification, then to be entitled to claim to have verified a statement cannot be to be entitled to claim a conclusive, *indefeasible* warrant for its assertion; for the most painstaking and careful execution of a computation confers no guarantee that it is correct." This is in effect a form of the familiar retreat (in respect of what warrants one's assertion), from "I have proved that it is so" (which, if true, surely equips one with a conclusive, indefeasible warrant) to "I have before me what, on painstaking and careful inspection, appears to be a proof that it is so." To suppose that this retreat is required is to miss the significance of the fact that if I am misled in such a case, the fault is in my moves in the space of reasons, not in the world. I suppose it is because Wright thinks mathematical proof and empirical verification are on a par in respect of vulnerability to the Argument from Illusion (and so in respect of the defeasibility of available warrants) that he thinks he can model empirical verification on mathematical verification without risking an undue concession to scepticism. (In effect Wright is committed to withholding, in respect of empirical verification, the acknowledgement that in n. 2 above I described as nearly universal.) I think the epistemology of empirical knowledge that results is disastrous; I try to bring this out in my "Mathematical Platonism and Dummettian Anti-Realism," *Dialectica* XLIII (1989), pp. 173–92.
- 5 It is important not to assume that, in rejecting this response, I am making unavailable the common-sense thought that we sometimes know how things are by seeing how they are. That would be so only if the epistemic status of such knowledge had to consist in the excellence of a policy or habit of basing belief on appearance, focused as it were on the particular case at hand. But that assumption is simply a form of what is under attack. (The status consists, rather, in the fact that one sees that things are so.)
- 6 On "blameless", see n. 2 above. However precisely it is spelled out, the idea of blameworthiness that we need most along with an idea of obligations as within one's power to discharge, on pain of losing contact with the point of interiorizing the space of reasons. So it is not to the point here to suggest that one can be blamed for a false belief based on appearance just because of its falsehood, on the analogy of the idea that one can be blamed for unintended consequences of one's intentional acts.
- 7 If a full-blown externalist approach preserves epistemological relevance for a concept of justification, it is certainly not one that functions as in the quotation from Sellars in n. 1 above.
- 8 For an example of the kind of hybrid account of knowledge that I am rejecting here, see Christopher Peacocke, *Thoughts: an Essay on Content* (Basil Blackwell, Oxford, 1986), chs 9 and 10, especially pp. 153–5. Peacocke has reliability as an external ingredient in a hybrid account of knowledge, one which also imposes an internal condition involving rationality. There is something similar in Simon Blackburn, "Knowledge, Truth, and Reliability," *Proceedings of the British Academy* LXX (1984), pp. 167–87, at pp. 178–9 (although on pp. 179–80

Blackburn comes close to acknowledging the internal importance of considerations of reliability).

9 Blackburn, "Knowledge, Truth, and Reliability," supposes that there is some deep error in insisting that a knower must be in an informational state which excludes all possibility that things are not as he takes them to be (a "guaranteeing" informational state). In reaction to that, he claims that titles to knowledge must be defensible "in the face of an open, acknowledged, possibility that the world might not be as we have come to take it to be" (p. 179). This may account for the fact that he seems not even to consider the truth requirement as an external condition on knowledge. But rejecting the idea that a knower's informational state is "guaranteeing" looks to me like rejecting a piece of plain common sense, that our locutions marking epistemically satisfactory positions ("see that . . .," "remember that . . .," and the like) are factive. That this is disastrous shows up in Blackburn's positive proposal, that one knows when no real possibility ("chance") that things are not as one takes them to be is left open by one's informational state, conceived as the upshot of the retreat supposedly forced by the Argument from Illusion – an "indicative" state, as opposed to a "guaranteeing" state. Blackburn applies this account of knowledge to the general hypotheses on which skeptical arguments trade (such as that one is a brain in a vat); the result is that whether one counts as knowing that such hypotheses do not obtain depends on who has the onus of proof in a dispute with a skeptic. But given how things look to someone on any particular occasion (and any other circumstances that might be relevant to reason's decision as to whether believing that things are that way is exercising a reliable policy or habit), there is surely a real possibility that things are not that way. That is just the point that (to put it in my terms) blamelessness in the space of reasons does not ensure factiveness in the position that results. In Blackburn's terms, misleading perceptual appearances, without surrounding clues to their misleadingness, are among "the kinds of things that happen" (p. 185). This point generalizes to other "indicative" states. So even if Blackburn achieves an onus-swapping standoff with the kind of skeptic who attempts to wield general skeptical hypotheses to undermine whole regions of knowledge all at once, his picture will deprive us of pretty much the same knowledge, only piece by piece. If we deny ourselves a "guaranteeing" conception of a putative knower's informational state, there will always be perfectly real possibilities (not the skeptic's arguably unreal possibilities) that he is wrong, given the lesser informational state we are committed to limiting him to. Blackburn simply misses this point; he concentrates on the general skeptical

hypotheses, as if there could be no threat to ordinary knowledge claims except from them.

10 When double-aspect views of content-involving mental states were a novelty, it used to be routine to cite the supposedly obvious compositeness of knowledge as an already familiar parallel. See, e.g., Daniel C. Dennett, "Beyond Belief," in Andrew Woodfield, ed., *Thought and Object* (Clarendon Press, Oxford, 1982), pp. 1–95, at pp. 11–12; and Colin McGinn, "The Structure of Content," *ibid.*, pp. 207–58, at p. 215. In querying the credentials of a hybrid conception of knowledge, I mean to do more than remove an expository prop from under those double-aspect views; I believe that direct extensions of the considerations in this lecture show that those views miss the point of the conceptual apparatus they aim to explain. I shall not be able to elaborate this here, though it will be close to the surface in §6.

11 As I noted in §2, we owe the world thanks for presenting us with appearances at all. But that point is accommodated by the formulation in the text. (If the world withheld appearances from us, reason would achieve its goal by deterring us from unsupported guesses as to how things are.)

12 Of course one can make mistakes; but the idea is that proper exercise of reason would eliminate them.

13 On the parallels in the sphere of practical reason, see Bernard Williams, "Moral Luck," in his *Moral Luck* (Cambridge University Press, Cambridge, 1981), pp. 20–39.

14 This formulation should make it clear how wildly off-target Blackburn is ("Knowledge, Truth, and Reliability," p. 176) in supposing that my appeal to "guaranteeing" informational states belongs within the general framework of the attempt "to ensure that there is no element of luck, or even contingency, in the true believer's title to knowledge." The traditional effect of this attempt to transcend luck is that the area of known fact is shrunk "potentially down to an entirely subjective realm." Blackburn takes me to offer a different option, within the same general framework, according to which, instead of a shrinkage in what can be known, the mind (the seat of these supposed luck-free "guaranteeing" states) expands to "embrace" all sorts of worldly states of affairs. This idea, which Blackburn rightly finds bizarre, has nothing to do with what I am proposing here, and was proposing in the work that Blackburn is discussing ("Criteria, Defeasibility, and Knowledge,"). Blackburn is so locked in to the framework of thought that epistemology must centre on a luck-free zone (a role played in his favored epistemology by the "indicative" states to which we are pushed back by the generalized Argument from Illusion) that he cannot comprehend how I can have been questioning the framework; so he has saddled me

- with the insane position that is the only interpretation my words will bear within the framework.
- 15 Exorcizing the fantasy should weaken the inclination to say that such a standing is not one's own achievement. Compare one of the practical analogues. The concept of what one does, understood as applying to one's interventions in the objective world, cannot mark out a sphere within which one has total control, immune to luck; only if we recoil from this into a fantasy of a sphere within which one's control is total does it seem to follow that what one genuinely achieves is less than one's interventions in the objective world. (This is one of many places at which much more discussion is needed.)
- 16 Seeing (or more generally perceiving) that things are a certain way is just one of the "factive" (or, in Blackburn's term, "guaranteeing") states that are restored to their proper status when the generalized Argument from Illusion is undermined; others include remembering how things were and learning from someone else how things are (see §2 above). In resisting the damaging effect of letting the Argument from Illusion structure epistemology (as in my "Criteria, Defeasibility, and Knowledge,"), I do not commit myself to assimilating all these "factive" positions to perception. Compare Crispin Wright, "Facts and Certainty," *Proceedings of the British Academy* LXXI (1985), 429–72. Commenting on that lecture of mine, he writes, at pp. 443–4: "Just as 'lifting' the veil of perception is to put us, on occasion anyway, in direct perceptual touch with material states of affairs, so a story has to be told explaining how we are similarly, on occasion, in direct perceptual touch with others' mental states and with past states of affairs – or at least, in direct perceptual touch with states of affairs which do better than provide inconclusive basis for claims about other minds and the past." I do think we are sometimes in direct perceptual touch with others' mental states, and certainly with states of affairs that do better than provide inconclusive basis for claims about them; but why should I accept the crazy idea that we are in direct *perceptual* touch with past states of affairs, when remembering them will plainly do instead? Similarly, when one learns something from someone else, the cognitive transaction is of course not a sort of perception of the state of affairs one is told about; resistance to letting the Argument from Illusion structure the epistemology of testimony need not involve denying that obvious fact (cf. Elizabeth Fricker, "The Epistemology of Testimony," *Aristotelian Society Supplementary Volume* LXI (1987), pp. 57–83, at pp. 74–5). I discuss the epistemology of testimony in "Knowledge by Hearsay," in B. K. Matilal and Arindam Chakrabarti, eds, *Knowing from Words* (Kluwer, Dordrecht, 1994), pp. 195–224.
- 17 When it turns out that the world has played us false, we conclude that it has presented us with a mere appearance rather than a manifest fact. Moreover, when the world does present us with a manifest fact, it does so by presenting us with an appearance. It is essential not to confuse these two pieces of common sense with the conclusion of the Argument from Illusion. Of course the content of the appearances that the world presents us with ("appearances" is here neutral as between "mere appearance" and "manifest fact") is not irrelevant to our possession of factive standings in the space of reasons. Our being able to count as, say, seeing that things are thus and so depends on our being properly sensitive (where "properly" expresses a rational assessment) to how things look to us. But it is a mistake to think this dependence is a matter of the appearance's functioning as a starting-point in the space of reasons, with the status of seeing how things are supposedly reconstructed in terms of a sufficiently cogent argument with the appearance as a premise. If the additional premises we can appeal to are restricted to what is available to reason on the interiorized conception of it, no such argument will be sufficiently cogent; that is a way of putting the reason why, once epistemology has started along the path marked out by the Argument from Illusion, the external supplementation is needed (§3 above). This is how we get into the position in which we have to choose between skepticism and the hybrid view. But the common-sense point that appearance bears on the rational status of belief is detachable from a commitment to that choice. (There is more discussion of this in my "Knowledge by Hearsay.")
- 18 Wright mentions two further reservations (besides the one I dealt with in n. 16 above) about my way with skepticism ("Facts and Certainty," p. 444). The first is this: "McDowell's proposal has... to be worked up into a demonstration that the sceptic actually has the epistemology of the various kinds of propositions *wrong*. The mere depiction of more comforting alternatives is not enough." I hope the present formulation of what I was trying to get at makes it clearer that this criticism misses the mark. My idea is that skepticism looks urgent only in the context of a visibly dubious assumption, which imposes a certain shape on the space of epistemological possibilities; so that the skeptic does indeed have the epistemology of the various kinds of propositions wrong. (But let me remark that my move is not well cast as an *answer* to skeptical challenges; it is more like a justification of a refusal to bother with them.) Wright's other reservation is that "'lifting' the veil of perception" has no obvious bearing on a style of skeptical argument exemplified by the attempt to undermine perceptual knowledge, or even perceptually grounded reasonable belief, on the basis that at any time at which one takes oneself

to have it, one lacks sufficient reason to believe that one is not dreaming. But I should have thought the bearing was quite obvious. Only if the veil is supposed to be in place can it seem that one would need to establish, or equip oneself with good reason to suppose, that one is not dreaming *before* one can be entitled to take one's apparent perceptions at face value. Once the veil is lifted, things can be the other way round; one's good reason to believe that one is not dreaming, on the relevant occasions, can reside in all the knowledge of the environment that one's senses are yielding one – something that does not happen when one is dreaming. (See my "Singular Thought and the Extent of Inner Space," in Philip Pettit and John McDowell (eds), *Subject, Thought, and Context* (Clarendon Press, Oxford, 1986), pp. 137–68, at pp. 147–8.)

- 19 I am quite unapologetic about the imagery here. Blackburn's *de haut en bas* remarks about my spatial imagery for the mental ("Knowledge, Truth, and Reliability," pp. 177–8) depend on a gross missing of its point; see n. 14 above.
- 20 Another element is the temptation to push all facts worthy of the name into an objective mould: the dualism results if we try to conceive subjectivity in an objectivistic way. I say more about this in "Singular Thought and the Extent of Inner Space," and in "Functionalism and Anomalous Monism," in Ernest LePore and Brian McLaughlin (eds), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Basil Blackwell, Oxford, 1986), pp. 387–98. A third element will emerge at the end of this lecture. (§8).
- 21 For an elaboration of this Sellarsian theme, see Richard Rorty, *Philosophy and the Mirror of Nature* (Basil Blackwell, Oxford, 1980).
- 22 I talk about this in "Singular Thought and the Extent of Inner Space."
- 23 Having invoked Sellarsian ideas (knowledge as a standing in the space of reasons; the rejection of the Given, or, what comes to the same thing, the rejection of a view of our conceptual scheme as what is set over against the Given; and the priority of reality over appearance), I ought to confess that I do not find in Sellars himself the direct figuring of manifest fact in the space of reasons that I am recommending. "Empiricism and the Philosophy of Mind" (in *Science, Perception and Reality*),

which is my source for all three of the ideas I have invoked, contains (in §35 and ff.) an account of the authority of observational reports which expresses a good thought – that the capacity to make observational reports requires general knowledge of the world, even in cases as conceptually undemanding as saying what color something is – in what seems to me a suspect way, in terms of the subject's ability to infer a judgment about the world from her own tokening (or propensity towards a tokening) of an observational form of words. I am suspicious of this avoidance of the straightforward idea that the authority of the report consists in the fact that things are manifestly so. (That idea is perfectly consistent with the good thought, not a relapse into a form of the Myth of the Given.) Sellars's account reflects some such idea as this: the content an expression has by virtue of its role in language-entry proprieties is non-conceptual content; conceptual content comes into play exclusively on the basis of inferential proprieties. I think this view of language-entry proprieties is a vestige of a bad way of thinking, which the main themes of Sellarsian philosophy show us how to undermine. But this paper is not meant as commentary on Sellars, and I shall not take these issues further here.

- 24 David Hume, *An Enquiry into the Human Understanding*, §4.
- 25 This is the basis of the denial, common in philosophy, that knowledge is a mental state: see, e.g., A. J. Ayer, *The Problem of Knowledge* (Penguin, Harmondsworth, 1956), pp. 14–26. (Cf. also n. 10 above.)
- 26 This is how it is with the mental states and occurrences that are most congenial to the notion of the mental as inner. The idea of the mental is complex, and can easily spread to cover states and occurrences of which this claim is not true. But these outliers do not trigger the philosophical moves that focus on the idea of interiority.
- 27 At §423 of *Philosophical Investigations* (Basil Blackwell, Oxford, 1953), the idea of things that go on in someone figures as a picture, one that Wittgenstein does not reject, though he suggests that there are difficulties in understanding its application. See also §427, on wanting to know what is going on in someone's head: "The picture should be taken seriously." (By "seriously" he does not mean "literally.")

Knowledge and the Social Articulation of the Space of Reasons

Robert Brandom

In "Knowledge and the Internal" (this volume, chapter 32) John McDowell presents a deep and interesting argument. I think everything he says is true and important. (Actually, I wouldn't want that quantifier to be restricted to the claims he makes there; I'm prepared to make this endorsement quite generally – but I won't try to defend *that* attitude here.) Still, there are a number of points that bear expanding on in order to be properly understood. So I want to say something about his point of departure: the idea of standings in the space of reasons. And I want to fill in further the picture at which he finally arrives, by saying how I think we *ought* to understand knowledge as a standing in the space of reasons, once we have freed ourselves from a prevalent deformed conception of that space. McDowell's strategy is to show that that conception of the space of reasons is inadequate – that it deserves to be called a "deformation" – by showing that it leaves no room for anything recognizable as knowledge. I'll try to reconstruct that argument by showing what it looks like in the context of a crucial dimension of the space of reasons that McDowell never mentions: its essentially *social* articulation. The effect of this supplementation, I think, is not to turn a bad argument into a good one, but to turn what is already a good argument into one that further illuminates the phenomena with which it deals.

The result may be just another confirmation of the maxim advising us that we ought to be more suspicious of philosophers who think they agree

with us than we are of philosophers who think that they don't. But depending on how successful I am, I hope at *worst* to clarify some of the key concepts and connections that McDowell appeals to, and at *best* to twist his words into a perverted caricature of their intended meaning. The game is worth the candle; for if I read him aright, when all the background he presupposes is made explicit, what he has offered us is nothing less than a generalized argument against all possible forms of epistemological externalism.

I

We can start with the Sellarsian idea that concepts are places in the space of reasons. According to this thought, to talk about the contents of beliefs and claims (the kinds of thing that are candidates for being or expressing knowledge) is to talk about things that can in principle be given as reasons, and for which reasons can in principle be asked.¹ One might think that in giving pride of place in this way to *justification* – by following Sellars in focussing to begin with on issues of what is a good reason for what – McDowell is begging important questions in the context of an investigation of what *knowledge* consists in. After all, it is the hallmark of the school of reliabilist externalism in contemporary epistemology precisely to deny that considerations of justification or what is a reason for what *need* have anything to do with assessments of knowledge. What distinguishes true beliefs that deserve to be called "knowledge" from those that

Originally published in *Philosophy and Phenomenological Research* 55 (1995), pp. 895–908.

do not is just that they have not been arrived at haphazardly or accidentally (for instance, by coin-flipping). What matters is that they be the outcome of a *reliable* belief-forming mechanism – one whose output is *likely* to be true. Under the right circumstances, making appropriate inferential moves in the space of reasons can qualify as such a mechanism. But it has no privileged status: knowledge can be diagnosed quite apart from any consideration of the space of reasons.

Epistemological externalism of this extreme sort, which regards the space of reasons as an optional superstructure, is one of the four positions on the nature of knowledge that McDowell considers. Even though his main interest lies elsewhere (in the more moderate externalism that sees considerations lying outside the space of reasons as only one element in a hybrid view), it might be thought that his dismissal of extreme externalism with the scornful remark that according to such an approach there is no principled reason not to count thermometers as knowers is a bit cavalier. In fact this is just the right thing to say, and no important questions about knowledge are being begged here. Seeing why this is so will help us to see better the role that is being played by the Sellarsian notion of a space of reasons.

For the important point has nothing to do with what one thinks of the propriety of the traditional construal of knowledge as *justified* true belief. It has to do with how one distinguishes concept use from nonconceptual activity. What is the difference between a parrot who is disposed reliably to respond differentially to the presence of red things by saying “Raawk, that’s red,” and a human reporter who makes the same noise under the same circumstances?² Or between a thermometer that responds to the temperature’s dropping below 70 degrees by reporting that fact by moving the needle on its output dial and a human reporter who makes a suitable noise under the same conditions? By hypothesis both reliably respond to the same stimuli, but we want to say that humans do, and the parrots and thermometers do not, respond by applying the *concepts* *red* or *70 degrees*. The parrot and the thermometer do not grasp those concepts, and so do not *understand* what they are “saying.” That is why we ought not to consider their responses as expressing *beliefs*: the *belief* condition on knowledge implicitly contains an *understanding* condition.

The Sellarsian idea with which McDowell begins is that this difference ought to be under-

stood in terms of the space of reasons. The difference that makes a difference in these cases is that for the human reporters, the claims “That’s red,” or “It’s 70 degrees out,” occupy positions in the space of reasons – the genuine reporters can tell what follows from them and what would be evidence for them. This practical know-how – being able to tell what they would be reasons for and what would be reasons for them – is as much a part of their understanding of “red” and “70 degrees” as are their reliable differential responsive dispositions. And it is this *inferential articulation* of those responses, the role they play in reasoning, that makes those responsive dispositions dispositions to apply *concepts*. If this idea is right, then nothing that can’t move in the space of reasons – nothing that can’t distinguish some claims or beliefs as *justifying* or being reasons for others – can even count as a concept user or believer, never mind a knower: it would be in another line of work altogether. And this point is not touched at all by the important observation that something that *is* in this line of work – something that *can* use concepts and have beliefs, something, that is, that can find its way around the space of reasons – can count as having knowledge in particular cases in which it has a true belief that it is not in a position to give reasons for. Extreme, or as I will henceforth feel entitled to call them, *gonzo* externalists mistakenly infer from the fact that issues of justification and reason-giving can be treated as *locally* irrelevant to attributions of knowledge in such cases, that they can safely be treated as *globally* irrelevant. The problem with this form of externalism is not with its construal of the *justification* condition on knowledge, but with its construal of the *belief* condition on knowledge.

These same terms give us some clues as to how we might think about the notion of *standings* in the space of reasons. A typical twenty-month-old child who toddles into the livingroom and in bell-like tones utters the sentence “The house is on fire,” is doing something quite different from what his seven-year-old sister would be doing by making the same noises. The young child is not claiming that the house is on fire, for the simple reason that he does not know what he would be committing himself to by that claim, what he would be making himself responsible for. He does not know what follows from it, what would be evidence for it, what would be incompatible with it, and so on. He does not know his way around the space of reasons well enough yet for anything he does to

count as adopting a standing in that space. His older sister knows that it follows from her claim that the family is in danger and should flee, and that the kitchen's being full of smoke and flame is evidence for it. She can commit herself, and knows what she would be committing herself to and what might entitle her to that commitment. She has begun to master the *inferential* articulation of such potential positions, statuses, or standings that make up the space of reasons – the things that can stand in the relation *is a reason for* to each other.

In order to clarify McDowell's argument and its conclusion, I'm going to recast them in an idiom he does *not* use: I'll talk about standings or statuses in the space of reasons in terms of two fundamental categories: *commitments* of a certain kind, and *entitlements* to those commitments. The idea is that occupying the basic sort of standing in the space of reasons is *staking a claim*, that is, undertaking a commitment of the sort that might be expressed by making a claim or assertion. Presystematically we might think of these as commitments to the truth of various propositions, that is, as beliefs. But I think it will be helpful if we keep talk of truth, propositions, and beliefs off-stage for a while. To uphold the fundamental Sellarsian idea about what would be required for these standings to have conceptual content, we must think about them as having two properties. First, it must be part of the conception of these commitments that the issue of one's *entitlement* to such a commitment can arise. Second, it must be possible for one such commitment to *inherit* or *derive* its entitlement from another. Together these mean that commitments can both serve as and stand in need of reasons. That is the sense in which they are being taken to be standings *in the space of reasons*.

The final point I want to make about McDowell's Sellarsian starting-point is that the "space of reasons" that he discusses ought to be understood as an abstraction from concrete practices of giving and asking for reasons. The space of reasons is a normative space. It is articulated by proprieties that govern practices of citing one standing as committing or entitling one to another – that is, as a reason for another. What people actually *do* is adopt, assess, and attribute such standings – and if they did not, there would be no such standings. For in the absence of such normative attitudes of taking or treating people as committed or entitled, there are no commitments or entitlements. They are not part of the furniture of the prehuman world.

When we turn to consider McDowell's diagnosis of deformations in our conception of the space of reasons that threaten to make knowledge unintelligible, it will be useful to keep our eyes on the actual practices of giving and asking for reasons, the practices that give a point to the abstract notion of a space of reasons.

II

McDowell's argument is structured by a botanization that classifies approaches to knowledge as coming in four flavors: skeptical, dogmatic, hybrid, and extreme externalist. We can group these further according to whether they conceive justification and truth as internally or externally related – or as I will say, according to whether they *aggregate* or *segregate* these conditions. The skeptic and the dogmatist take it as a criterion of adequacy on a notion of justification that any claim or belief that is sufficiently justified is true. They are right that if a claim or belief has the status of knowledge, it is guaranteed to be true. But they also take it that justification of a certain sort is what distinguishes knowledge from other belief. If that is right, then justification must be truth-guaranteeing. While *agreeing* on this basic principle, the skeptic and the dogmatist *disagree* about whether a notion of justification meeting this condition is to be had. The skeptic arrives at the *false* conclusion that knowledge is *not* possible by combining the *false* claim that justification must be incompatible with falsehood with the *true* claim that justification that rules out the possibility of falsehood is *not* to be had. The dogmatist arrives at the *true* conclusion that knowledge *is* possible by combining the *false* claim that justification must be incompatible with falsehood with the further *false* claim that justification that rules out the possibility of falsehood *can* be had. McDowell rightly does not rehearse at length the difficulties of these views; their unsatisfactoriness is widely acknowledged.

Where skepticism and dogmatism run the justification and truth conditions on knowledge too closely together, the hybrid and extreme versions of externalism drive them too far apart. Gonzo externalism throws justification and the giving of reasons out entirely, and I've already indicated why this won't do.³ It is by no means obvious, though, why the more moderate externalism of a two-factor or hybrid view can't be made to work. McDowell's core argument is accordingly devoted

to showing what is wrong with them. The danger he sees is that if a satisfactory standing with respect to justification or reason-giving is seen as an internal matter, something one can secure all on one's own, while assessments of truth or reliability answer to external standards, then the justification and truth conditions on knowledge are treated as independent of one another. But being justified in holding a belief just is being justified in taking it to be *true*. Such segregationist views, he argues, are inherently unstable and untenable.

Although his four-part classification lumps them together, McDowell implicitly acknowledges two different forms hybrid views can take. They deserve to be considered together because each *extrudes* from the space of reasons some sort of assessment that is crucial to the attribution of knowledge, segregating it as external to standings in that space. The first version extrudes both considerations of truth and of reliability as statuses distinguished from that of being justified. That is, assessments of what is *true*, and of the *reliability* of various policies for endorsing some claim as true (believing it), are taken to be independent in principle from questions of what is a reason for what. McDowell rightly gives this sort of segregationism short shrift: nothing recognizable as our notion of justification survives if our justificatory practices are forbidden in principle from being criticized and shaped on the basis of assessments of their reliability, that is, the likelihood that reasoning in the ways they sanction will lead to truths. This recognition of the intimate connection between justification and reliability motivates the second version of hybrid segregationism, which seeks to incorporate assessments of reliability into standings in the space of reasons, while still extruding truth. But since the reliability of a belief-endorsing policy just is the likelihood that it will lead to the endorsement of truths, this version of the hybrid approach is no more stable a position than the other.

The underlying fact is that the notions of belief, justification, reliability, and truth are inextricably intertwined in ways that preclude the segregationist approach. I started by talking about how the notion of belief, as a conceptually contentful state, is unintelligible apart from considerations of what is a reason for what – that is, apart from liability to assessments of *justification*. But it is equally essential to our notion of beliefs that they are something for which the question of *truth* can arise: believing is taking or treating as *true*. Any adequate account

of the contentfulness of beliefs and claims must show why and how these two crucial dimensions of assessment are so intimately linked. As McDowell's argument indicates, a useful way to unpack that linkage is by looking at the concept of *reliability*. For, on the one hand, assessment of cognitive reliability makes sense only against a background that includes assessments of truth: a belief-endorsing policy is reliable just insofar as it is likely to lead to truths. And on the other hand, assessments of justification must answer to assessments of reliability. Arguing that a proposed method of justification is not likely to lead to truths is not just *one* way of criticizing such methods – any more than stopping the heart is just one way of killing a vertebrate; it is the form all those ways share, the common conclusion they must lead to if they are to be successful. That is why McDowell can put such pressure on segregationist accounts of knowledge by focusing attention on the notion of reliability: the hybrid approaches are unstable because they can neither adequately construe the space of reasons independently of considerations of reliability, nor adequately construe reliability assessments apart from truth assessments.

We can think of McDowell's argument as coming in three nested parts. The core argument is the one I've just rehearsed, which uses the notion of reliability to underscore that considerations of truth cannot be extruded from the space of reasons. This argument contributes the crucial premise in a wider argument, to the effect that *none* of the four currently available approaches to knowledge is satisfactory: neither skepticism, nor dogmatism, nor hybrid theories, nor extreme externalism. The segregationist approaches drive the justification and truth conditions on knowledge too far apart – in the case of gonzo externalism, at the cost of losing sight of what distinguishes beliefs and claims as conceptually articulated, by ignoring the space of reasons altogether. But the two aggregationist approaches are equally unsatisfactory, running justification and truth together so that their distinctive contributions to knowledge assessments are confounded. The third part of McDowell's argument is the claim that these four approaches share a presupposition concerning the shape of the space of reasons, and that given that presupposition, they exhaust the alternatives. If that is right, then the unsatisfactoriness of these ways of construing knowledge shows that that presupposition must be

rejected, and the space of reasons otherwise understood.

III

I said at the outset that I think this is a good argument. I've now indicated how I think the first two steps work: the argument for the instability of hybrid views and, based on it, the argument for the unsatisfactoriness of any of the four kinds of epistemological segregationism and aggregationism he distinguishes. Two large issues remain: the diagnosis of these approaches as generated by a shared erroneous conception of the space of reasons, and the recommendation of an alternative. It is at this point that I would like to offer what I regard (though McDowell may not) as a friendly amendment to or clarification of his account.

Early on in his paper, McDowell gives us the following characterization of the conceptual pathology he takes to have generated the shuttling back and forth between unsatisfactory positions (whether aggregationist or segregationist) characteristic of contemporary and classical epistemology:

The deformation is an interiorization of the space of reasons, a withdrawal of it from the external world. This happens when we suppose that we ought to be able to achieve flawless standings in the space of reasons by our own unaided resources, without needing the world to do us any favors.

Now I don't want to disagree with this, but I do want to insist that this diagnosis should come at the end of a story, not at the beginning. You may have noticed that although here and there I helped myself to McDowell's imagery of what is conceived as internal or external to the space of reasons, in my exposition of his core arguments I did not find it necessary to say *anything at all* about *interiorizing* the space of reasons in this sense – did not need to talk at all about what candidate knowers are supposed to be able to do *all on their own* as opposed to what they can only do by grace of *favors from the world*. The deformed and defective conception of the space of reasons that I see as underlying the various forms of epistemological aggregationism and segregationism in play so far is prior to, and, I want to argue, explanatory of, the one McDowell focuses on.

For I want to claim that the mistake is to begin with to *individualize* the space of reasons. The complaint I want to make about McDowell's discussion is that he makes nothing of the essential *social* articulation of that space. The passage above is typical: he says the thought is that *we* ought to be able to achieve flawless standings in the space of reasons by *our* own resources, without needing the world to do *us* any favors; but for all he says here or elsewhere, this *us* could be *each* of us, individually or by ourselves, rather than *all* of us collectively. But this difference makes all the difference.

The best way I know to make this clear is to try to indicate in more detail than McDowell does just what an account looks like that *does* construe factive statuses such as knowledge as standings in the space of reasons, while respecting the lessons of externalism. I said above that it is important to remember that our abstract talk about reasons and the space of reasons has to be grounded in an appreciation of the concrete practices of giving and asking for reasons, namely in what people actually *do*. I also suggested that what McDowell calls *standings* in the space of reasons should be thought of in terms of *commitments* and *entitlements* that are practically acknowledged by those engaging in such practices. What I want to claim now is that if we recognize that giving and asking for reasons is a constellation of essentially *social* practices, and that the commitments and entitlements those practices involve are accordingly essentially *social* statuses, we will be in a position to understand factive locutions such as *knowledge* and warrantive locutions such as *reliable* as attributing standings in the space of reasons. Furthermore, we will be able to make appropriate sense of the different roles of assessments of truth and of justification in attributions of knowledge, as the aggregationist approaches of the skeptic and the dogmatist could not, without disjoining those roles so severely as to engender the problems we saw with the various segregationist approaches, in particular the embarrassment that hybrid theories have concerning the notion of reliability.

The key to understanding knowledge as a standing in the space of reasons is to focus on the practical attitude adopted by one who is assessing a candidate for such a standing: What is someone who attributes knowledge *doing*? For these purposes we can continue to be guided, as we have throughout, by the traditional conception of knowledge as justified true belief. Construed as a standing or status, belief will correspond to some

sort of commitment, while justification (being justified) will correspond to some sort of entitlement to that commitment. So taking someone to have a justified belief will be understood as attributing two sorts of standings: a commitment and an entitlement.

What about the truth condition? To take someone to have the status of a knower one must take it that the justified belief in question is also *true*. What is it to do that? Taking the belief in question to be true is not a matter of *attributing* a commitment, but of *undertaking* one – endorsing the claim oneself. For taking-true is just believing, that is, committing oneself, adopting a standing or status.⁴ What sort of case leads us to distinguish justified beliefs that are true from those that are not? If you are standing in a darkened room and seem to see a candle ten feet in front of you, I may take you to have good reason for believing that there is a candle ten feet in front of you, and so to take you to be entitled to your commitment. But that may be my attitude even if I know, as you do not, that there is a mirror five feet in front of you, and no candle behind it, so that I am not in a position to endorse or commit myself to what you are committed to.

Thinking of things this way, assessing someone as having successfully achieved the status or standing of a knower involves adopting three different attitudes: *attributing* a commitment, *attributing* an entitlement, and *undertaking* a commitment. There is nothing in principle mysterious about such assessments, nor, therefore, about the standing being assessed. Knowledge is intelligible as a standing in the space of reasons, because and insofar as it is intelligible as a status one can be taken to achieve in the game of giving and asking for reasons. But it is essentially a *social* status, because it incorporates and depends on the *social* difference of perspective between *attributing* a commitment (to another) and *undertaking* a commitment (oneself). If one *individualizes* the space of reasons, forgetting that it is a *shared* space within which we adopt attitudes towards *each other* – and so does not think about standings in the space of reasons as socially articulated, as potentially including the social difference of perspective between attributing and undertaking commitments, that is, between your standing and mine – then one will not be able to understand knowledge as a standing in the space of reasons. One will then have either to try to get some individualized standing to do the work of the socially articulated factive, as the

aggregationists do, or to extrude some components of it from the space of reasons entirely, as the segregationists do. One will then be doomed either to lose the crucial distinction between the belief and justification conditions, on the one hand, and the truth condition on knowledge, on the other, as the aggregationists do, or to lose the crucial connections between them, as the segregationists do.

The distinction of social perspective between attributing a standing and adopting it keeps the truth condition from being run together with the others, and so makes it possible to understand assessments of something as having the standing of a justified belief that is not true. But McDowell's core argument indicates that the danger of distinguishing these elements too firmly – losing the crucial connections – manifests itself in difficulties with notion of *reliability*. What can we say about this test case?

A fundamental point on which broadly externalist approaches to epistemology are clearly right is that one can *be* justified without being *able* to justify. That is, one can have the standing of being *entitled* to a commitment without having to *inherit* that entitlement from *other* commitments inferentially related to it as reasons. A paradigm case is that of noninferential reports. If you are a generally reliable noninferential reporter of lighted candles in darkened rooms, then you can be entitled to your claim that there is a candle in front of you in cases where that claim or commitment arose by your exercise of that reliable differential disposition to respond to such candles by making such reports (undertaking such commitments, adopting such standings). And this can be the case even if you are not able yourself to cite your reliability in such matters as a reason for the belief you acquired.⁵

Now it would be wrong to conclude from the fact that a piece of knowledge can be acquired noninferentially even where the knower is unable to justify it that reasons need not be in play at all. For to begin with, you must be capable of making the claim or acquiring the belief in order to be a candidate for knowing it. And that requires that you *understand* it: that you have at least a rough practical mastery of its inferential role, the know-how to discriminate some things that follow from it from others that don't, and some things that would be evidence for it from others that would not. In "Empiricism and the Philosophy of Mind" Sellars unfortunately takes it that in order to secure *this* claim, he must insist that one is not

justified unless one *knows* one is justified – in particular, that noninferential reports should be accorded the status of knowledge only in cases where the knower can cite her own reliability as a reason, from which the correctness of the noninferential report could be inferred. This response is excessive; there is no reason to deny the externalist insight that, once one is capable of achieving standings in the space of reasons – for instance capable of committing oneself to the claim that there is a candle in the room – one can become entitled to such standings without being able to give reasons for them. But Sellars's overreaction also contains an important insight: reliability matters to assessments of knowledge precisely because of the *inferences* it can support.

The key point to understanding reliability as a warrantive standing in the space of reasons is that the notion of reliability itself is essentially an *inferential* notion: a matter precisely of what is a reason for what. What must be kept in mind if one is to talk (a variant of) the traditional language of justification as *internal* entitling and reliability as *external* entitling is that what they are internal or external *to* is not the practice of giving and asking for reasons, and so not the space of reasons, but rather the individual whose standings in that space are being assessed. For reliability is precisely a matter of a *socially* articulated inference. For me to take you to be a reliable reporter of lighted candles in darkened rooms is just for me to endorse a particular pattern of reasoning; in particular it is for me to endorse the inference that could be made explicit by saying:

If in a darkened room S noninferentially acquires the belief that there is a lighted candle, *then* (probably) there is a lighted candle there.

Translated into the language I have suggested for discussing standings in the space of reasons (that is, statuses one can acquire in the game of giving and asking for reasons) this is an inferential connection between a suitably noninferentially acquired commitment *attributed* to you and a corresponding commitment that I *undertake*. It is treating your commitment as a (defeasible) reason for my own.

The externalist epistemologist who takes reliability to warrant the attribution of noninferential knowledge in the absence of justification relies precisely on this essentially *interpersonal* pattern of inference. If we like, we can say that Sellars's

point is reinstated at one remove of social perspective: although it is enough that the subject of knowledge *be* reliable to be entitled to a belief (without having to be able to *cite* that reliability as a reason for it), the *attributor* of knowledge must be able to cite that reliability as such a reason.⁶

My conclusion is that if we keep firmly in mind that the space of reasons is founded on practices of giving and asking for reasons – practices in which standings or statuses can not only be adopted but attributed – then we can understand truth and reliability, no less than justification, and hence in the end even *knowledge*, as socially articulated standings in that space. According to such a picture, the insights of externalism are accommodated as pointing to features of the essentially social practices of giving and asking for reasons, undertaking and attributing inferentially articulated commitments and entitlements. So issues of justification, on the one hand, and of truth and reliability, on the other, are not severed from one another, and the instability that McDowell diagnoses in what he calls “hybrid” views is avoided. Knowledge and reliability, involving as they do essential reference to truth, *are* in a certain sense *hybrid* statuses on this account. For they are made intelligible by appeal to *two* different social perspectives, that of the one to whom a status is attributed and that of the one attributing it.⁷ But the crucial difference is that this is a distinction of perspectives *within* the space of reasons, not a distinction between what is within it and what is without it. That is why the instability McDowell points to does not arise.

IV

My response to McDowell's paper has come in two parts, one constructive, and one critical. Constructively, I have indicated how knowledge can be construed as a standing in the space of reasons. The key question is what I must be *doing* in order to take you to have that standing. And the answer is, in line with the JTB account of knowledge, that corresponding to the belief condition (which includes an understanding condition) I must *attribute* a propositionally contentful commitment, that corresponding to the justification condition I must *attribute* also entitlement to that commitment (whether inferentially or noninferentially grounded), and that corresponding to the truth condition I must also myself *endorse* or *undertake*

the same propositionally contentful commitment. The possibility of such *truth* assessments is already implicit in the attribution of a commitment corresponding to belief – for propositionally contentful commitments are essentially, and not merely accidentally, things for which the question of truth can arise. Likewise, and for that reason, the possibility of *reliability* assessments is already implicit in the attribution of an entitlement corresponding to the justification condition on knowledge. But assessing your reliability is a matter of whether to endorse a certain pattern of interpersonal inference: the inference from a commitment I *attribute* to you to one I *undertake* myself. In this way what is expressed by the use of *factive* locutions such as “believes truly,” *warrantive* locutions, such as “believes reliably,” and *cognitive* locutions such as “knows,” which include both factive and warrantive dimensions, can all be understood as standings in a socially articulated space of reasons: standings that incorporate what are with respect to *individual* knowers internal and external epistemic considerations in the form of the distinct *social* perspectives of attributing and undertaking commitments.

It is the different perspectives provided by different sets of commitments that make it possible to *triangulate* on objective states of affairs. Our practices of comparing, assessing, and correcting different repertoires of commitments one with respect to another – those we attribute to others and those we undertake ourselves – are what make them intelligible as *perspectives*, *views of* something, ways in which a perspective-independent reality can *appear*.⁸ To individualize the space of reasons is to interiorize it. To ignore the social articulation of standings in the space of reasons is

to leave out what makes it possible to understand such standings as answerable for their correctness to how things actually are. And such an interiorized rendering must in the end fail, as McDowell insists, even to be recognizable as *belief*. For what an individualized construal leaves out is what makes statuses such as knowledge and reliability intelligible as standings in the space of reasons. Factives, like *believes truly* (because of their relation to truth-assessment) warrantives, like *believes justifiedly* (because of their relation to reliability), and so cognitives, like *knows*, testify at once to the way in which objective facts (concerning how things really are, not just how they are taken to be) are incorporated in the space of reasons, and equally how the social articulation of that space makes such incorporation so much as intelligible.

My only complaints against McDowell have been accusations of sins of omission – a matter of what he has *not* said. Such complaints are often unfair: one can't say everything. But I have not reproached him for saying nothing about the effects of the discovery of silver in the New World on the spread of the Hussite heresy in Central Europe, even though he has indeed been silent regarding this important topic. I have reproached him for saying nothing about the social articulation of the space of reasons in the context of a discussion of a deformed conception of the space of reasons that makes it impossible for us to understand how knowledge and reliability are related to such standings. For that disastrous interiorizing of the space of reasons results precisely from individualizing it.

So let me end as I began: everything McDowell says is true and important – but sometimes he leaves stuff out.

Notes

1 As I use the terms, what can serve as premises and conclusions of inferences are *propositional* contents, which is the fundamental and defining species of *conceptual* contents. Conceptual contents that are not propositional correspond to what is expressed by subsentential expressions, and are to be understood by abstraction from the propositional contents of sentences containing them – which is to say that the contribution a subsentential expression makes to the propositional contents expressed by sentences in which it occurs is to be identified by observing the effects on those contents of substituting other expressions for it.

2 Perhaps minus the parrot expletive.

3 I said above that McDowell offers us a generalized argument against any possible form of epistemological externalism. (It might better be thought of as a recipe that, given any such externalist view, shows us how to construct a knock-down argument against it.) In conversation, McDowell points to one case that might be thought to be an exception. Non- or pre-linguistic animals do not have status or standing in the space of reasons. So according to the idiom being recommended here, they neither deploy concepts, acquire beliefs, nor count as having knowledge. Nonetheless, it is common to talk about them loosely

as though they were capable of some version (usually admitted to be degenerate cases) of these accomplishments. The informational states most closely resembling genuine beliefs that they *do* have (call them *beliefs**), when they both correctly represent how things are and are acquired by a suitable reliable process may be called *knowledge**. An externalist account of *this* sort of state is all that is to be had. This status has in common with the genuine article what the parrot has in common with the reporter of red things: reliable differential responsive dispositions.

- 4 The root notion of truth is just what the tradition always took it to be: saying of what is that it is. The cases in which I take it that *p*, that is where I believe or am committed to the claim that *p*, are just the cases in which I take your belief that *p* to be true. The mistake of metaphysical conceptions of truth, including any substantive correspondence theory, is to assimilate what I am doing when I take your belief to be true to what I am doing when I take you to believe it, or to be justified in doing so. For in those cases I *attribute* a commitment and an entitlement, respectively. Metaphysical theories of truth are theories of the property I am attributing to your commitment when I take it to be true. But in taking it to be true I am not attributing *any* property to that commitment, I am endorsing it myself.
- 5 Where reliabilist counterexamples undermine the *necessity* of the JTB account of knowledge, Gettier-style counterexamples undermine its *sufficiency*. Justifications that depend essentially on false premises (even ones the candidate knower is entitled to believe) may be assessed as inadequate support for attributions of knowledge. Although the point cannot be pursued here, this phenomenon can be accommodated in the social-perspectival framework presented here, by looking at the relation between what would be reasons for the one to whom a candidate piece of knowledge is attributed, on the one hand, and what would be reasons for the one who attributes them (from whose point of view the truth of those premises is assessed).
- 6 Assuming sufficient expressive resources are available to formulate the reliability inference explicitly in a conditional (which makes it available in a form suitable to serve as a premise in further inferences).
- 7 When I talk about “the social articulation of the space of reasons” I mean that standings in that space must be understood in terms of these two kinds of socially related perspective. I do *not* mean that the community is privileged in some way relative to individuals. So it would be a mistake to think of my remarks as suggesting a super-individual sort of interiorization, in which the community as a whole plays the role formerly played by particular individuals.
- 8 I develop this view, and the other constructive suggestions offered in the second half of this paper, in greater detail in *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Harvard University Press, 1994).

PART VIII

Virtue Epistemology and Proper Cognitive Functioning

Introduction

Virtue epistemology and proper functionalism may be fairly regarded as descendants of reliabilism. Both views require, for knowledge, that one's belief be produced by a reliable process, but both deny this is sufficient. Thus, virtue theory requires, at least, that the belief-producing process also be an *ability*, a stable disposition to acquire or maintain beliefs, and proper functionalism requires that it be a faculty functioning the way it was designed to function. These restrictive modifications of reliabilism are aimed in part at solving some of the well-known problems facing reliabilism, including what John Greco calls the problem of epistemic responsibility, viz., the problem posed by reliably produced beliefs that fall short of knowledge (justification) owing to the epistemic irresponsibility of the believer, and the new evil demon problem, viz., the problem of accounting for the justification possessed by victims of the evil demon, subjects whose faculties are unreliable.

In his contribution, Alvin Goldman outlines tasks and methods for epistemology. Epistemology has a descriptive and a normative task. The descriptive is to characterize our commonsense epistemic concepts and norms, our "folkways," that is, *how* we go about evaluating beliefs epistemically. The normative task, which depends on the descriptive, is to formulate a set of norms that improves upon the folkway norms. In the selection in this section, Goldman focuses on the descriptive task. His resulting view is that ordinary folk epistemically evaluate in the following way. Working with a devised, or socially inherited, list of epistemic virtues and vices, one considers actual and hypothetical cases of beliefs. Any such belief is counted as justified if its corresponding process is

similar to a process on the list of virtues, as unjustified if its process is similar to a process on the list of vices, and as non-justified otherwise. The spirit of reliabilism is not lost in this account, for the processes on the list of virtues are there because we have deemed them reliable.

Applied to the problem of epistemic responsibility and the new evil demon problem, the method delivers the following results, according to Goldman. The beliefs of hypothetical reliable clairvoyants, for example, are non-justified, if clairvoyance is dissimilar from both our recognized virtues and vices, or unjustified, if clairvoyance is similar to one of our recognized vices. Since the victims of the evil demon use our most trusted processes, the beliefs formed using these processes are counted as justified.

To the question "Why wouldn't we simply add clairvoyance to our list of virtues if we found it reliable?" the reply is that we are categorially conservative. Our presumption is always against revising our list of virtues and vices. Goldman finds confirmation for this claim in the work of empirical psychologists.

In his selection, Alvin Plantinga outlines proper functionalism. Proper functionalism is a theory of *warrant*, viz., of that feature that turns true belief into knowledge. The account Plantinga provides, notably, works with the normative notion of a faculty's functioning *properly*, and thus is a departure from standard externalist accounts. To function properly a faculty must function as it ought to function, that is, as it is designed to function. Our cognitive design plan, importantly, has a segment devoted to the production of true beliefs. A belief is warranted, then, only if the faculty that

Introduction

produced it was aiming at truth. But this is not enough. The faculty must be reliable, perhaps not *tout court*, but in the kinds of environment for which the subject's faculties were designed. A human clairvoyant in an imagined situation in which clairvoyance is reliable employs a belief-forming process that hasn't been given a role in the design plan, and so his clairvoyant beliefs, however reliable, lack warrant. Although victims of an evil demon lack full-blown warrant on Plantinga's account, one can see that there is room for assigning their beliefs some degree of positive epistemic status, since their faculties are functioning properly and would reliably produce true beliefs in the kinds of environments for which they were designed.

Linda Zagzebski and John Greco ask the question of how best to conceive of epistemic virtue. Zagzebski claims that a virtue, whether moral or intellectual, is an excellence, acquired through time and work, that contains both a motivational element and a success element. Correlated with every virtue is a motivation directed on some goal, and no ability can be a virtue unless its possessor succeeds in achieving that goal. The goal unifying the intellectual virtues is the understanding of reality. Reliable success at achieving this goal, moreover, isn't limited to the overall production of more truths than falsehoods. A virtue, such as originality or intellectual courage, may be reliable in the further sense of being an ability the possession of which helps to make possible the advancement of understanding in a domain of inquiry. As long as such traits, working together with other virtues, operate to correct errors produced along the way, they serve the ultimate epistemic goal of understanding the world.

John Greco argues that in order to solve the problems facing virtue theories, internalist ele-

ments must be introduced. It is not enough to say, with Sosa and Goldman, that the victims of the evil demon are justified because their faculties would be reliable in our environment. Even supposing our faculties are unreliable, their beliefs are justified. Facts about the success of our faculties are simply irrelevant to the matter of *their* justification. Nor can the problem of epistemic responsibility be solved by adding the requirement of a perspective on one's ways of believing. To achieve the right results, perspectives must be specified in a complicated way that the ordinary knower cannot be expected to understand.

The problems can be solved, however, if we introduce the internalist notion of a subject's conformance to a norm she countenances. To countenance a norm is to be guided by it in conscientious reasoning. To be in conformance with norms one countenances is not merely to believe in accordance with them, but to believe *because* one countenances them. Epistemically irresponsible believers either flout norms they countenance or fail to be in conformance with them. Greco gives the example of a person, Mary, who unwittingly possesses a special non-experiential device for detecting tigers. Mary's belief formed using this device cannot count as knowledge, since in forming the belief, she has flouted norms she countenances, for among the norms she countenances are norms that forbid forming beliefs about the presence of tigers in the absence of any evidence to that effect. In connection with the victims of the evil demon, Greco argues that, although they lack warrant, they have justification, for their beliefs are in conformance with norms they countenance. This, moreover, is a feature they possess independently of facts about the reliability of our faculties.

Further Reading

- Axtell, Guy, "Recent Work on Virtue Epistemology," *American Philosophical Quarterly* 34 (1997), pp. 1–26.
- Code, Lorraine, *Epistemic Responsibility* (Hanover, NH: University Press of New England, 1987).
- Greco, John, "Internalism and Epistemically Responsible Belief," *Synthese* 85 (1990), pp. 245–77.
- , "Virtues and Vices of Virtue Epistemology," *Canadian Journal of Philosophy* 23 (1993), pp. 413–32.

- Kvanvig, Jonathan L., *The Intellectual Virtues and the Life of the Mind* (Lanham: Rowman & Littlefield, 1992).
- Kvanvig, Jonathan L. (ed.), *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge* (Lanham, MD: Rowman and Littlefield, 1996).
- Montmarquet, James A., *Epistemic Virtue and Doxastic Responsibility* (Lanham, MD: Rowman and Littlefield, 1993).

Plantinga, Alvin, *Warrant and Proper Function* (Oxford: Oxford University Press, 1993).

Sosa, Ernest, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991), Parts II–IV.

Zagzebski, Linda Trinkhaus, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press, 1996).

Epistemic Folkways and Scientific Epistemology

Alvin I. Goldman

I

What is the mission of epistemology, and what is its proper methodology? Such meta-epistemological questions have been prominent in recent years, especially with the emergence of various brands of "naturalistic" epistemology. In this paper, I shall reformulate and expand upon my own meta-epistemological conception (most fully articulated in Goldman 1986), retaining many of its former ingredients while reconfiguring others. The discussion is by no means confined, though, to the meta-epistemological level. New substantive proposals will also be advanced and defended.

Let us begin, however, at the meta-epistemological level, by asking what role should be played in epistemology by our ordinary epistemic concepts and principles. By some philosophers' lights, the sole mission of epistemology is to elucidate commonsense epistemic concepts and principles: concepts like knowledge, justification, and rationality, and principles associated with these concepts. By other philosophers' lights, this is not even part of epistemology's aim. Ordinary concepts and principles, the latter would argue, are fundamentally naive, unsystematic, and uninformed by important bodies of logic and/or mathematics. Ordinary principles and practices, for example, ignore or violate the probability calculus, which ought to be the cornerstone of epistemic rationality. Thus, on the second view, proper epistemology must neither end with naive principles of justification or rationality, nor even begin there.

My own stance on this issue lies somewhere between these extremes. To facilitate discussion, let us give a label to our commonsense epistemic concepts and norms; let us call them our *epistemic folkways*. In partial agreement with the first view sketched above, I would hold that *one* proper task of epistemology is to elucidate our epistemic folkways. Whatever else epistemology might proceed to do, it should at least have its roots in the concepts and practices of the folk. If these roots are utterly rejected and abandoned; by what rights would the new discipline call itself "epistemology" at all? It may well be desirable to reform or transcend our epistemic folkways, as the second of the views sketched above recommends. But it is essential to preserve continuity; and continuity can only be recognized if we have a satisfactory characterization of our epistemic folkways. Actually, even if one rejects the plea for continuity, a description of our epistemic folkways is in order. How would one know what to criticize, or what needs to be transcended, in the absence of such a description? So a first mission of epistemology is to describe or characterize our folkways.

Now a suitable description of these folk concepts, I believe, is likely to depend on insights from cognitive science. Indeed, identification of the semantic contours of many (if not all) concepts can profit from theoretical and empirical work in psychology and linguistics. For this reason, the task of describing or elucidating folk epistemology is a *scientific* task, at least a task that should be informed by relevant scientific research.

The second mission of epistemology, as suggested by the second view above, is the formulation of a more adequate, sound, or systematic set of

Originally published in A. Goldman, *Laurens* (Cambridge, MA: MIT Press, 1992), pp. 155-63.

epistemic norms, in some way(s) transcending our naive epistemic repertoire. How and why these folkways might be transcended, or improved upon, remains to be specified. This will partly depend on the contours of the commonsense standards that emerge from the first mission. On my view, epistemic concepts like knowledge and justification crucially invoke psychological faculties or processes. Our folk understanding, however, has a limited and tenuous grasp of the processes available to the cognitive agent. Thus, one important respect in which epistemic folkways should be transcended is by incorporating a more detailed and empirically based depiction of psychological mechanisms. Here too epistemology would seek assistance from cognitive science.

Since both missions of epistemology just delineated lean in important respects on the deliverances of science, specifically cognitive science, let us call our conception of epistemology *scientific epistemology*. Scientific epistemology, we have seen, has two branches: *descriptive* and *normative*. While descriptive scientific epistemology aims to describe our ordinary epistemic assessments, normative scientific epistemology continues the practice of making epistemic judgments, or formulating systematic principles for such judgments.¹ It is prepared to depart from our ordinary epistemic judgments, however, if and when that proves advisable.

II

Mainstream epistemology has concentrated much of its attention on two concepts (or terms): knowledge and justified belief. This essay focuses on the latter. We need not mark this concept exclusively by the phrase "justified belief." A family of phrases pick out roughly the same concept: "well-founded belief," "reasonable belief," "belief based on good grounds," and so forth. I shall propose an account of this concept that is in the reliabilist tradition, but departs at a crucial juncture from other versions of reliabilism. My account has the same core idea as Ernest Sosa's *intellectual virtues* approach, but incorporates some distinctive features that improve its prospects.²

The basic approach is, roughly, to identify the concept of justified belief with the concept of belief obtained through the exercise of intellectual virtues (excellences). Beliefs acquired (or retained) through a chain of "virtuous" psychological pro-

cesses qualify as justified; those acquired partly by cognitive "vices" are derogated as unjustified. This, as I say, is a *rough* account. To explain it more fully, I need to say things about the psychology of the epistemic evaluator, the possessor and deployer of the concept in question. At this stage in the development of semantical theory (which, in the future, may well be viewed as part of the "dark ages" of the subject), it is difficult to say just what the relationship is between the meaning or "content" of concepts and the form or structure of their mental representation. In the present case, however, I believe that an account of the form of representation can contribute to our understanding of the content, although I am unable to formulate these matters in a theoretically satisfying fashion.

The hypothesis I wish to advance is that the epistemic evaluator has a mentally stored set, or list, of cognitive virtues and vices. When asked to evaluate an actual or hypothetical case of belief, the evaluator considers the processes by which the belief was produced, and matches these against his list of virtues and vices. If the processes match virtues only, the belief is classified as justified. If the processes are matched partly with vices, the belief is categorized as unjustified. If a belief-forming scenario is described that features a process not on the evaluator's list of either virtues or vices, the belief may be categorized as neither justified nor unjustified, but simply *nonjustified*. Alternatively (and this alternative plays an important role in my story), the evaluator's judgment may depend on the (judged) *similarity* of the novel process to the stored virtues and vices. In other words, the "matches" in question need not be perfect.

This proposal makes two important points of contact with going theories in the psychology of concepts. First, it has some affinity to the *exemplar* approach to concept representation (cf. Medin and Schaffer 1978; Smith and Medin 1981; Hintzman 1986). According to that approach, a concept is mentally represented by means of representations of its positive instances, or perhaps types of instances. For example, the representation of the concept *pants* might include a representation of a particular pair of faded blue jeans and/or a representation of the type *blue jeans*. Our approach to the concept of justification shares the spirit of this approach insofar as it posits a set of examples of virtues and vices, as opposed to a mere abstract characterization – e.g., a definition – of

(intellectual) virtue or vice. A second affinity to the exemplar approach is in the appeal to a similarity, or matching, operation in the classification of new target cases. According to the exemplar approach, targets are categorized as a function of their similarity to the positive exemplars (and dissimilarity to the foils). Of course, similarity is invoked in many other approaches to concept deployment as well (see E. E. Smith 1990). This makes our account of justification consonant with the psychological literature generally, whether or not it meshes specifically with the exemplar approach.

Let us now see what this hypothesis predicts for a variety of cases. To apply it, we need to make some assumptions about the lists of virtues and vices that typical evaluators mentally store. I shall assume that the virtues include belief formation based on sight, hearing, memory, reasoning in certain "approved" ways, and so forth. The vices include intellectual processes like forming beliefs by guesswork, wishful thinking, and ignoring contrary evidence. *Why* these items are placed in their respective categories remains to be explained. As indicated, I plan to explain them by reference to reliability. Since the account will therefore be, at bottom, a reliabilist type of account, it is instructive to see how it fares when applied to well-known problem cases for standard versions of reliabilism.

Consider first the demon-world case. In a certain possible world, a Cartesian demon gives people deceptive visual experiences, which systematically lead to false beliefs. Are these vision-based beliefs justified? Intuitively, they are. The demon's victims are presented with the same sorts of visual experiences that we are, and they use the same processes to produce corresponding beliefs. For most epistemic evaluators, this seems sufficient to induce the judgment that the victims' beliefs are justified. Does our account predict this result? Certainly it does. The account predicts that an epistemic evaluator will match the victims' vision-based processes to one (or more) of the items on his list of intellectual virtues, and therefore judge the victims' beliefs to be justified.

Turn next to Laurence Bonjour's (1985) cases in which hypothetical agents are assumed to possess a perfectly reliable clairvoyant faculty. Although these agents form their beliefs by this reliable faculty, Bonjour contends that the beliefs are not justified; and apparently most (philosophical) evaluators agree with that judgment. This result is not predicted by simple forms of reliabilism.³ What does our present theory predict? Let

us consider the four cases in two groups. In the first three cases (Samantha, Casper, and Maud), the agent has contrary evidence that he or she ignores. Samantha has a massive amount of apparently cogent evidence that the president is in Washington, but she nonetheless believes (through clairvoyance) that the president is in New York City. Casper and Maud each has large amounts of ostensibly cogent evidence that he/she has no reliable clairvoyant power, but they rely on such a power nonetheless. Here our theory predicts that the evaluator will match these agent's belief-forming processes to the vice of ignoring contrary evidence. Since the processes include a vice, the beliefs will be judged to be unjustified.

Bonjour's fourth case involves Norman, who has a reliable clairvoyant power but no reasons for or against the thesis that he possesses it. When he believes, through clairvoyance, that the president is in New York City, while possessing no (other) relevant evidence, how should this belief be judged? My own assessment is less clear in this case than the other three cases. I am tempted to say that Norman's belief is *non*justified, not that it is thoroughly *un*justified. (I construe unjustified as "having negative justificational status," and nonjustified as "lacking positive justificational status.") This result is also readily predicted by our theory. On the assumption that I (and other evaluators) do not have clairvoyance on my list of virtues, the theory allows the prediction that the belief would be judged neither justified nor unjustified, merely nonjustified. For those evaluators who would judge Norman's belief to be *un*justified, there is another possible explanation in terms of the theory. There is a class of putative faculties, including mental telepathy, ESP, telekinesis, and so forth that are scientifically disreputable. It is plausible that evaluators view any process of basing beliefs on the supposed deliverances of such faculties as vices. It is also plausible that these evaluators judge the process of basing one's belief on clairvoyance to be *similar* to such vices. Thus, the theory would predict that they would view a belief acquired in this way as unjustified.⁴

Finally, consider Alvin Plantinga's (1988) examples that feature disease-triggered or mind-malfunctioning processes. These include processes engendered by a brain tumor, radiation-caused processes, and the like. In each case Plantinga imagines that the process is reliable, but reports that we would not judge it to be justification-conferring. My diagnosis follows the track outlined in

the Norman case. At a minimum, the processes imagined by Plantinga fail to match any virtue on a typical evaluator's list. So the beliefs are at least nonjustified. Furthermore, evaluators may have a prior representation of pathological processes as examples of cognitive vices. Plantinga's cases might be judged (relevantly) similar to these vices, so that the beliefs they produce would be declared unjustified.

In some of Plantinga's cases, it is further supposed that the hypothetical agent possesses countervailing evidence against his belief, which he steadfastly ignores. As noted earlier, this added element would strengthen a judgment of unjustifiedness according to our theory, because ignoring contrary evidence is an intellectual vice. Once again, then, our theory's predictions conform with reported judgments.

Let us now turn to the question of how epistemic evaluators acquire their lists of virtues and vices. What is the basis for their classification? As already indicated, my answer invokes the notion of reliability. Belief-forming processes based on vision, hearing, memory, and ("good") reasoning are deemed virtuous because they (are deemed to) produce a high ratio of true beliefs. Processes like guessing, wishful thinking, and ignoring contrary evidence are deemed vicious because they (are deemed to) produce a low ratio of true beliefs.

We need not assume that each epistemic evaluator chooses his/her catalogue of virtues and vices by direct application of the reliability test. Epistemic evaluators may partly inherit their lists of virtues and vices from other speakers in the linguistic community. Nonetheless, the hypothesis is that the selection of virtues and vices rests, ultimately, on assessments of reliability.

It is not assumed, of course, that all speakers have the same lists of intellectual virtues and vices. They may have different opinions about the reliability of processes, and therefore differ in their respective lists.⁵ Or they may belong to different subcultures in the linguistic community, which may differentially influence their lists. Philosophers sometimes seem to assume great uniformity in epistemic judgments. This assumption may stem from the fact that it is mostly the judgments of philosophers themselves that have been reported, and they are members of a fairly homogeneous subculture. A wider pool of "subjects" might reveal a much lower degree of uniformity. That would conform to the present theory, however, which permits individual differences in cata-

logues of virtues and vices, and hence in judgments of justifiedness.

If virtues and vices are selected on the basis of reliability and unreliability, respectively, why doesn't a hypothetical case introducing a novel reliable process induce an evaluator to add that process to his list of virtues, and declare the resulting belief justified? Why, for example, doesn't he add clairvoyance to his list of virtues, and rule Norman's beliefs to be justified?

I venture the following explanation. First, people seem to have a trait of *categorical conservatism*. They display a preference for "entrenched" categories, in Nelson Goodman's (1955) phraseology, and do not lightly supplement or revise their categorical schemes. An isolated single case is not enough. More specifically, merely imaginary cases do not exert much influence on categorical structures. People's cognitive systems are responsive to live cases, not purely fictional ones. Philosophers encounter this when their students or nonphilosophers are unimpressed with science fiction-style counterexamples. Philosophers become impatient with this response because they presume that possible cases are on a par (for counterexample purposes) with actual ones. This phenomenon testifies, however, to a psychological propensity to take an invidious attitude toward purely imaginary cases.

To the philosopher, it seems both natural and inevitable to take hypothetical cases seriously, and if necessary to restrict one's conclusions about them to specified "possible worlds." Thus, the philosopher might be inclined to hold, "If reliability is the standard of intellectual virtue, shouldn't we say that clairvoyance is a virtue *in the possible worlds* of Bonjour's examples, if not a virtue in general?" This is a natural thing for philosophers to say, given their schooling, but there is no evidence that this is how people naturally think about the matter. There is no evidence that "the folk" are inclined to relativize virtues and vices to this or that possible world.

I suspect that concerted investigation (not undertaken here) would uncover ample evidence of conservatism, specifically in the normative realm. In many traditional cultures, for example, loyalty to family and friends is treated as a cardinal virtue.⁶ This view of loyalty tends to persist even through changes in social and organizational climate, which undermine the value of unqualified loyalty. Members of such cultures, I suspect, would continue to view personal loyalty as a virtue

even in *hypothetical* cases where the trait has stipulated unfortunate consequences.

In a slightly different vein, it is common for both critics and advocates of reliabilism to call attention to the relativity of reliability to the domain or circumstances in which the process is used. The question is therefore raised, what is the relevant domain for judging the reliability of a process? A critic like John Pollock (1986, pp. 118–19), for example, observes that color vision is reliable on earth but unreliable in the universe at large. In determining the reliability of color vision, he asks, which domain should be invoked? Finding no satisfactory reply to this question, Pollock takes this as a serious difficulty for reliabilism. Similarly, Sosa (1988 and 1991) notes that an intellectual structure or disposition can be reliable with respect to one field of propositions but unreliable with respect to another, and reliable in one environment but unreliable in another. He does not view this as a difficulty for reliabilism, but concludes that any talk of intellectual virtue must be relativized to field and environment.

Neither of these conclusions seems apt, however, for purposes of *description* of our epistemic folkways. It would be a mistake to suppose that ordinary epistemic evaluators are sensitive to these issues. It is likely – or at least plausible – that our ordinary apprehension of the intellectual virtues is rough, unsystematic, and insensitive to any theoretical desirability of relativization to domain or environment. Thus, as long as we are engaged in the description of our epistemic folkways, it is no criticism of the account that it fails to explain what domain or environment is to be used. Nor is it appropriate for the account to introduce relativization where there is no evidence of relativization on the part of the folk.

Of course, we do need an explanatory story of how the folk arrive at their selected virtues and vices. And this presumably requires some reference to the domain in which reliability is judged. However, there may not be much more to the story than the fact that people determine reliability scores from the cases they personally “observe.” Alternatively, they *may* regard the observed cases as a sample from which they infer a truth ratio in some wider class of cases. It is doubtful, however, that they have any precise conception of the wider class. They probably don’t address this theoretical issue, and don’t do (or think) anything that commits them to any particular resolution of it. It would therefore be wrong to expect descriptive

epistemology to be fully specific on this dimension.

A similar point holds for the question of process individuation. It is quite possible that the folk do not have highly principled methods for individuating cognitive processes, for “slicing up” virtues and vices. If that is right, it is a mistake to insist that descriptive epistemology uncover such methods. It is no flaw in reliabilism, considered as descriptive epistemology, that it fails to unearth them. It may well be desirable to develop sharper individuation principles for purposes of normative epistemology. But the missions and requirements of descriptive and normative epistemology must be kept distinct.

This discussion has assumed throughout that the folk have lists of intellectual virtues and vices. What is the evidence for this? In the moral sphere ordinary language is rich in virtues terminology. By contrast, there are few common labels for intellectual virtues, and those that do exist – “perceptiveness,” “thoroughness,” “insightfulness,” and so forth – are of limited value in the present context. I propose to identify the relevant intellectual virtues (at least those relevant to *justification*) with the belief-forming capacities, faculties, or processes that would be accepted as answers to the question “How does *X* know?” In answer to this form of question, it is common to reply, “He saw it,” “He heard it,” “He remembers it,” “He infers it from such-and-such evidence,” and so forth. Thus, basing belief on seeing, hearing, memory, and (good) inference are in the collection of what the folk regard as intellectual virtues. Consider, for contrast, how anomalous it is to answer the question “How does *X* know?” with “By guesswork,” “By wishful thinking,” or “By ignoring contrary evidence.” This indicates that *these* modes of belief formation – guessing, wishful thinking, ignoring contrary evidence – are standardly regarded as intellectual *vices*. They are not ways of obtaining knowledge, nor ways of obtaining justified belief.

Why appeal to “knowledge”-talk rather than “justification”-talk to identify the virtues? Because “know” has a greater frequency of occurrence than “justified,” yet the two are closely related. Roughly, justified belief is belief acquired by means of the same sorts of capacities, faculties, or processes that yield knowledge in favorable circumstances (i.e., when the resulting belief is true and there are no Gettier complications, or no relevant alternatives).

To sum up the present theory, let me emphasize that it depicts justificational evaluation as involving two stages. The first stage features the acquisition by an evaluator of some set of intellectual virtues and vices. This is where reliability enters the picture. In the second stage, the evaluator applies his list of virtues and vices to decide the epistemic status of targeted beliefs. At this stage, there is no direct consideration of reliability.

There is an obvious analogy here to rule utilitarianism in the moral sphere. Another analogy worth mentioning is Saul Kripke's (1980) theory of *reference-fixing*. According to Kripke, we can use one property to fix a reference to a certain entity, or type of entity; but once this reference has been fixed, that property may cease to play a role in identifying the entity across various possible worlds. For example, we can fix a reference to heat as the phenomenon that causes certain sensations in people. Once heat has been so picked out, this property is no longer needed, or relied upon, in identifying heat. A phenomenon can count as

heat in another possible world where it doesn't cause those sensations in people. Similarly, I am proposing, we initially use reliability as a test for intellectual quality (virtue or vice status). Once the quality of a faculty or process has been determined, however, it tends to retain that status in our thinking. At any rate, it isn't reassessed each time we consider a fresh case, especially a purely imaginary and bizarre case like the demon world. Nor is quality relativized to each possible world or environment.

The present version of the virtues theory appears to be a successful variant of reliabilism, capable of accounting for most, if not all, of the most prominent counterexamples to earlier variants of reliabilism.⁷ The present approach also makes an innovation in naturalistic epistemology. Whereas earlier naturalistic epistemologists have focused exclusively on the psychology of the epistemic agent, the present paper also highlights the psychology of the epistemic evaluator.

Notes

- 1 Normative scientific epistemology corresponds to what I elsewhere call *epistemics* (see Goldman 1986). Although epistemics is not restricted to the assessment of *psychological* processes, that is the topic of the present paper. So we are here dealing with what I call *primary epistemics*.
- 2 Sosa's approach is spelled out most fully in Sosa 1985, 1988, and 1991.
- 3 My own previous formulations of reliabilism have not been so simple. Both "What Is Justified Belief?" (Goldman 1979) and *Epistemology and Cognition* (Goldman 1986) had provisions – e.g., the non-undermining provision of *Epistemology and Cognition* – that could help accommodate Bonjour's examples. It is not entirely clear, however, how well these qualifications succeeded with the Norman case, described below.
- 4 Tom Senor presented the following example to his philosophy class at the University of Arkansas. Norman is working at his desk when out of the blue he is hit (via clairvoyance) with a very distinct and vivid impression of the president at the Empire State Building. The image is phenomenally distinct from a regular visual impression but is in some respects similar and of roughly equal force. The experience is so overwhelming that Norman just can't help but form the belief that the president is in New York. About half of Senor's class judged that in this case Norman justifiably believes that the president is in New York. Senor points out, in commenting on this paper, that their judgments are readily explained by the present account, because the description of the clairvoyance process makes it sufficiently similar to vision to be easily "matched" to that virtue.
- 5 Since some of these opinions may be true and others false, people's lists of virtues and vices may have varying degrees of accuracy. The "real" status of a trait as a virtue or vice is independent of people's opinions about that trait. However, since the enterprise of descriptive epistemology is to describe and explain evaluators' judgments, we need to advert to the traits they *believe* to be virtues or vices, i.e., the ones on their mental lists.
- 6 Thanks to Holly Smith for this example. She cites Riding 1989 (ch. 6) for relevant discussion.
- 7 It should be noted that this theory of justification is intended to capture the *strong* conception of justification. The complementary conception of *weak* justification will receive attention elsewhere.

References

- BonJour, Laurence, 1985. *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press).
- Goldman, Alvin I., 1979. "What Is Justified Belief?" this vol., ch. 27.
- , 1986. *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- Goodman, Nelson, 1955. *Fact, Fiction and Forecast* (Cambridge, MA: Harvard University Press).
- Hintzman, D., 1986. "'Schema Abstraction' in a Multiple-Trace Memory Model," *Psychological Review* 93, pp. 411–28.
- Kripke, Saul, 1980. *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- Medin, D. L. and M. M. Schaffer, 1978. "A Context Theory of Classification Learning," *Psychological Review* 85, pp. 207–38.
- Plantinga, Alvin, 1988. "Positive Epistemic Status and Proper Function," in J. Tomberlin (ed.), *Philosophical Perspectives*, vol. 2 (Atascadero, CA: Ridgeview).
- Pollock, John, 1986. *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield).
- Riding, A., 1989. *Distant Neighbors: A Portrait of the Mexicans* (New York: Vintage Books).
- Smith, E. E., 1990. "Categorization," in D. Osherson and E. Smith (eds), *Thinking: An Invitation to Cognitive Science* (Cambridge, MA: MIT Press).
- Smith, E. E. and M. Medin, 1981. *Categories and Concepts* (Cambridge, MA: Harvard University Press).
- Sosa, Ernest, 1985. "Knowledge and Intellectual Virtue," *The Monist* 68, pp. 226–63.
- , 1988. "Beyond Scepticism, to the Best of Our Knowledge," *Mind* 97, pp. 153–88.
- , 1991. "Reliabilism and Intellectual Virtue," in *Knowledge in Perspective* (Cambridge: Cambridge University Press).

Warrant: A First Approximation

Alvin Plantinga

One thought emerging from our canvas of contemporary accounts of warrant in *Warrant: The Current Debate* is that there are many different valuable epistemic states of affairs – epistemic values, we might call them, giving that oft-abused word a decent sense; and different conceptions of warrant appeal to different epistemic values. For example, there is doing one's subjective epistemic duty, doing one's objective epistemic duty, and doing both; these figure prominently in classical internalism. There is having a set of beliefs that is coherent to one or another degree; there is also the *disposition* to have coherent beliefs; these things are what the coherentist is quite naturally enthusiastic about. There is having adequate evidence or good reasons for your beliefs, this goes with the evidentialism that has been a dominant feature of the epistemological tradition and is presently represented in different ways by Feldman and Conee,¹ and William Alston.² There is having a reliable set of faculties or belief-producing mechanisms, which of course goes with reliabilism of various sorts. There is also knowing that you have a reliable set of epistemic faculties. There is also Foley rationality; and there are the several varieties of Foley rationality, such as believing what you *think* would contribute to your attaining your epistemic goal, believing what on reflection you *would* think would contribute to your attaining that goal, believing what *really would* contribute to your doing so, and so on. There is having a set of beliefs that contributes to your nonepistemic goals such as

happiness, or living the good life, or living the moral life. There is having the *right* goals; there is *aiming* to have the right goals; and there is *knowing* that you have the right goals. There is believing what is true, and there is having true beliefs on important topics; there is accepting a given belief to the right degree. There is knowing that you know; there is being able to prove to the skeptic that you know. And there are a thousand other epistemic virtues.

I Proper Function

Now the notion of warrant is clearly connected with all of these epistemic values and more besides. (The problem here is to come up with a conception of warrant that gives to each its due and describes how each is connected with the others and with warrant.) As a first step toward developing a satisfying account of warrant, I should like to call attention to still another epistemic value: having epistemic faculties that *function properly*. The first thing to see, I think, is that this notion of proper function is the rock on which the canvassed accounts of warrant founder. Cognitive malfunction has been a sort of recurring theme. Chisholm's dutiful epistemic agent who, whenever he is appeared to redly, always believes that nothing is appearing redly to him, Pollock's cognizer who by virtue of malfunction has the wrong epistemic norms, the Coherent but Inflexible Climber, Dretske's epistemic agent whose belief that Spot emits ultraviolet radiation has been caused by the fact that Spot does indeed emit such radiation, Goldman's victim of the

Originally published in A. Plantinga, *Warrant and Proper Function* (Oxford and New York: Oxford University Press, 1993), pp. 3–20.

epistemically serendipitous lesion: all are such that their beliefs lack warrant for them. In each case the reason, I suggest, is *cognitive malfunction*, failure of the relevant cognitive faculties to function properly, to function as they ought to. Chisholm's agent meets Chisholm's conditions for warrant; his beliefs lack warrant, however, because they result from cognitive dysfunction due to a damaging brain lesion, or the machinations of an Alpha Centaurian scientist, or perhaps the mischievous schemes of a Cartesian evil demon. Something similar must be said for each of the others. In each case the unfortunate in question meets the conditions laid down for warrant by the account in question; in each case her beliefs fail to have warrant because of cognitive malfunction. Hence each of these accounts misfires, at least in part by virtue of its failure to take appropriate account of the notion of proper function.

I therefore suggest initially that a necessary condition of a belief's having warrant for me is that my cognitive equipment, my belief-forming and belief-maintaining apparatus or powers, be free of such malfunction. A belief has warrant for you only if your cognitive apparatus is functioning properly, working the way it ought to work, in producing and sustaining it. (Of course this isn't nearly sufficient, and I shall try to supply some of what is necessary to achieve sufficiency.)

The notion of proper function is one member of a connected group of interdefinable notions; some of the other members of the group are *dysfunction*, *design*, *function* (simpliciter), *normality* (in the normative nonstatistical sense), *damage*, and *purpose*. There is initial reason to doubt, I think, that this circle of concepts can be broken into from the outside – that is, reason to doubt that any of them can be defined without reference to the others. Here we have a situation like that with modality: possibility, contingency, necessity, entailment, and their colleagues form a circle of properties or concepts that can be defined or explained in terms of each other but cannot be defined in terms of properties outside the circle. (Of course that is nothing against these modal concepts.) The same goes here, I think.

You may nonetheless think there is a serious problem with this notion right from the start. Isn't the idea of proper function an extremely unlikely idea to appeal to in explaining the notion of warrant? Isn't it every bit as puzzling, every bit as much in need of explanation and clarification, as the notion of warrant itself? Perhaps so; but even if

so, at least we can reduce our total puzzlement by explaining the one in terms of the other; and we can see more clearly the source and location of some of our perplexities about warrant. Further, the idea of proper function is one we all have; we all grasp it in at least a preliminary rough-and-ready way; we all constantly employ it. You go to the doctor; he tells you that your thyroid isn't functioning quite as it ought (its thyroxin output is low); he prescribes a synthetic thyroxin. If you develop cataracts, the lenses of your eyes become less transparent; they can't function properly and you can't see well. A loss in elasticity of the heart muscle can lead to left ventricular malfunction. If a bird's wing is broken, it typically won't function properly; the bird won't be able to fly until the wing is healed, and then only if it heals in such a way as not to inhibit proper function. Alcohol and drugs can interfere with the proper function of various cognitive capacities, so that you can't drive properly, can't do simple addition problems, display poor social judgment, get into a fist fight, and wind up in jail.

And it isn't just in rough-and-ready everyday commonsense contexts that the notion of proper function is important; it is deeply embedded in science.

We are accustomed to hearing about biological functions for various bodily organs. The heart, the kidneys, and the pituitary gland, we are told, have functions – things they are, in this sense *supposed to do*. The fact that these organs are supposed to do these things, the fact that they have their functions, is quite independent of what *we* think they are supposed to do. Biologists *discovered* these functions; they didn't invent or assign them. We cannot, by agreeing among ourselves, *change* the functions of these organs. . . . The same seems true for sensory systems, those organs by means of which highly sensitive and continuous dependencies are maintained between external, public events and internal, neural processes. Can there be a serious question about whether, in the same sense in which it is the heart's function to pump the blood, it is, say, the task or function of the noctuid moth's auditory system to detect the whereabouts and movements of its archenemy, the bat?³

According to David Baltimore, "many instances of blood disorders, mental problems, and a host of

other disabilities are traceable to a malfunctioning gene.”⁴ According to the great Swiss child psychologist Jean Piaget, a seven-year-old child whose cognitive faculties are functioning properly will believe that everything in the universe has a purpose in some grand overarching plan or design; later on a properly functioning person, he said, will learn to “think scientifically” and realize that everything has either a natural cause or happens by chance.⁵

Biological and social scientists, furthermore – psychologists, medical researchers, neuroscientists, economists, sociologists, and many others – continually give accounts of how human beings or other organisms or their parts and organs function: how they work, what their purposes are, and how they react under various circumstances. Call these descriptions (following John Pollock)⁶ *functional generalizations*. For example, whenever a person is appeared to redly under such and such conditions, she will form the belief that there is something red present; whenever a person considers an obvious *a priori* truth such as $2 + 1 = 3$, she will find herself firmly believing it; whenever a person desires something and believes so and so, he will do such and such. To strike a more sophisticated if no more enlightening note: whenever an organism of kind K is in state S_i and receives sensory input P_i , then there is a probability of r that it will go into state S_j and produce output O_j . Pollock makes the important point that if these functional generalizations are taken straightforwardly and at face value, as universal generalizations about people and other organisms and their parts, they are nearly always false. They don’t hold of someone who is in a coma, having a stroke, crazed by strong drink, or has just hit the ground after a fall off a cliff. Clearly these functional generalizations contain something like an implicit restriction to organisms and organs that are *functioning properly*, functioning as they ought to, subject to no malfunction or dysfunction. The notion of proper function, therefore, is presupposed by the idea of functional generalizations.

So the notion of proper function is a notion we have and regularly employ; I may therefore appeal to it in explaining warrant. Still, it needs exploration, clarification, and explication if it is to serve as the key notion in an account of warrant. Let us provisionally entertain the idea that a belief has warrant for me only if the relevant parts of my noetic equipment – the parts involved in its formation and sustenance – are functioning properly.

It is easy to see, however, that proper function cannot be the whole story about warrant. You have just had your annual cognitive checkup at MIT; you pass with flying colors and are in splendid epistemic condition. Suddenly and without your knowledge you are transported to an environment wholly different from earth; you awake on a planet revolving around Alpha Centauri. There conditions are quite different; elephants, we may suppose, are invisible to human beings, but emit a sort of radiation unknown on earth, a sort of radiation that causes human beings to form the belief that a trumpet is sounding nearby. An Alpha Centaurian elephant wanders by; you are subjected to the radiation, and form the belief that a trumpet is sounding nearby. There is nothing wrong with your cognitive faculties; they are working quite properly; still, this belief has little by way of warrant for you. Nor is the problem merely that the belief is false; even if we add that a trumpet really is sounding nearby (in a soundproof telephone booth, perhaps), your belief will still have little by way of warrant for you.

To vary the example, imagine that the radiation emitted causes human beings to form the belief not that a trumpet is sounding, but that there is a large gray object in the neighborhood. Again, an elephant wanders by; while seeing nothing of any particular interest, you suddenly find yourself with the belief that there is a large gray object nearby. A bit perplexed at this discovery, you examine your surroundings more closely: you still see no large gray object. Your faculties are displaying no malfunction (you have your certificate from MIT); you are not being epistemically careless or slovenly (you are doing your epistemic best); nevertheless you don’t know that there is a large gray object nearby. That belief has little or no warrant for you. Of course you may be justified, within your epistemic rights in holding this belief; you may be flouting no epistemic duty. Further, the belief may also be rational for you in every sensible sense of “rational.”⁷ But it has little warrant for you.

What this example is designed to show, of course, is that the proper function of your epistemic equipment is not (logically) sufficient for warrant: it is possible that your cognitive equipment be functioning perfectly properly but your beliefs still lack warrant for you. And the reason is not far to seek: it is that your cognitive faculties and the environment in which you find yourself are not properly attuned. The problem is not with your cognitive faculties; they are in good working order.

The problem is with the environment – with your cognitive environment. In approximately the same way, your automobile might be in perfect working order, despite the fact that it will not run well at the top of Pike's Peak, or under water, or on the moon. We must therefore add another component to warrant; your faculties must be in good working order, and the environment must be appropriate for your particular repertoire of epistemic powers. It must be the sort of environment for which your faculties are designed – by God or evolution (or both). Perhaps there are creatures native to the planet in question who are much like human beings but whose cognitive powers fit that epistemic environment and differ from ours in such a way that Alpha Centaurian elephants are not invisible to them. Then their beliefs would have warrant where yours do not.

It is tempting to suggest that warrant *just* is (or supervenes upon) proper functioning in an appropriate environment, so that a given belief has warrant for you to the degree that your faculties are functioning properly (in producing and sustaining that belief) in an environment appropriate for your cognitive equipment: the better your faculties function, the more warrant. But this cannot be correct. Couldn't it happen that my cognitive faculties are working properly (in an appropriate environment) in producing and sustaining a certain belief in me, while nonetheless that belief enjoys less by way of warrant for me than some other belief? Say that a pair of beliefs are (for want of a better term) *productively equivalent* if they are produced by faculties functioning properly to the same degree and in environments of equal appropriateness. Then couldn't it be that a pair of my beliefs should be productively equivalent while nonetheless one of them has more by way of warrant – even a great deal more – than the other? Of course that could be; as a matter of fact it happens all the time. The belief that $7 + 5 = 12$, or the belief that I have a name, or the belief that I am more than seven years old – any of these has more by way of warrant for me than does the memory belief, now rather dim and indistinct, that forty years ago I owned a secondhand sixteen-gauge shotgun and a red bicycle with balloon tires; but all, I take it, are produced by cognitive faculties functioning properly in a congenial environment. Although both epistemic warrant and *being properly produced* come in degrees, there seems to be no discernible functional relationship between them: but then we can't see warrant as simply a matter of

a belief's being produced by faculties working properly in an appropriate environment. We still have no real answer to the question *What is warrant?* That particular frog (with apologies to John Austin) is still grinning residually up from the bottom of the mug.

Fortunately there is an easy response. Not only does the first belief, the belief that $7 + 5 = 12$, have more by way of warrant for me than the second; it is also one I accept much more firmly. It seems *obviously* true, in a way in which the belief about the bicycle and shotgun do not. Among the things we believe, we believe some much more firmly than others. I believe that I live in Indiana, that $2 + 1 = 3$, that the sun is larger than the earth, that China has a larger population than India, and that Friesland used to be much larger than it is now; and I believe some of these things more firmly than others. Here I speak of full belief, not the partial beliefs of which Bayesians speak.⁸ Following Ramsey, Bayesians sometimes suggest that my degrees of belief can be at least roughly determined by examining my betting behavior; the least odds at which I will bet on a proposition *A* measures the degree to which I believe *A*. If I am willing to bet at odds of 2:1 that the die will come up either 5 or 6 then I must believe to degree .667 that it will come up that way. This seems to me wrong. The truth is I believe it *probable* to degree .667 that the die will come up that way. And no doubt I fully believe *that*; that is, in this case I don't believe *anything* to degree .667 (strictly speaking, there is no such thing as believing something to degree .667), but I *do* believe (fully believe) that there is a .667 probability that the die will come up either 5 or 6. Suppose I buy a ticket in a thousand-ticket lottery I believe to be fair. Here it is false, I think, that I believe I will not win, or believe that to degree .999. What I do believe is that it is very *probable* (probable to degree .999) that I won't win.⁹

Return to the case in question, then: although I believe both $7 + 5 = 12$ and *40 years ago I owned a secondhand 16-gauge shotgun and a red bicycle with balloon tires*, I believe the former more strongly than the latter; this is correlated with the fact that the former has more by way of warrant for me than the latter. I therefore conjecture that when my cognitive establishment is working properly, then in the typical case, the degree to which I believe a given proposition will be proportional to the degree it has of warrant – or if the relationship

isn't one of straightforward proportionality, some appropriate functional relationship will hold between warrant and this impulse. When my faculties are functioning properly, a belief has warrant to the degree that I find myself inclined to accept it; and this (again, if my faculties are functioning properly and nothing interferes) will be the degree to which I *do* accept it.

Initially, and to (at most) a zeroeth approximation, therefore, we may put it like this: in the paradigm cases of warrant, a belief *B* has warrant for *S* if and only if that belief is produced in *S* by his epistemic faculties working properly in an appropriate environment; and if both *B* and *B** have warrant for *S*, *B* has more warrant than *B** for *S* iff *S* believes *B* more firmly than *B**. And knowledge requires both true belief, and a certain degree of warrant (a degree that may vary from context to context, so that knowledge may display a certain indexical character).¹⁰

Putting the matter thus imports what is at this stage at best a wholly spurious pretense of precision and completeness; and the rest of this chapter will be given over to some of the necessary qualifications, amplifications, and the like, including attention to the absolutely crucial notion of the design plan. To begin with some of the essential and obvious qualifications then: it is of first importance to see that this condition – that of one's cognitive equipment functioning *properly* – is not the same thing as one's cognitive equipment functioning *normally*, not, at any rate, if we take the term "normally" in a broadly statistical sense. Even if one of my systems functions in a way far from the statistical norm, it might still be functioning properly. (Alternatively, what we must see is that there is a distinction between a normative and statistical sense of "normal.") Carl Lewis is not defective with respect to jumping by virtue of the fact that he can jump much further than the average person. Perhaps most adult tomcats get into lots of fights and ordinarily move into late middle age with patches of fur torn out; it does not follow that an old tomcat with all of his fur suffers from some sort of tonsorial disorder. Perhaps most male cats get neutered; it does not follow that those that don't are incapable of proper function. If, by virtue of some nuclear disaster, we were nearly all left blind, it would not follow that the few sighted among us would have improperly functioning eyes. So your belief's being produced by your faculties working *normally* or in *normal* conditions – that is, the sorts of conditions that most fre-

quently obtain – must be distinguished from their working *properly*.

Further, a belief has warrant for me only if my epistemic faculties are working properly in producing and sustaining it; but of course it isn't true that *all* of my cognitive faculties have to be functioning properly in order for a given belief to have warrant for me. Suppose my memory plays me tricks; obviously that does not mean that I can't have warrant for such introspective propositions as that I am appeared to redly. What must be working properly are the faculties (or subfaculties, or modules) involved in the production of the particular belief in question. And even they need not be working properly over the entire range of their operation. Suppose I cannot properly hear high notes: I may still learn much by way of the hearing ability I do have. Furthermore, a faculty that does not function properly *without outside aid* can nonetheless furnish warrant; I can have warrant for visual propositions even if I need glasses and can see next to nothing without them. Still further, even if my corrected vision is very poor, I can still have warrant for visual propositions; even if I can't perceive colors at all, I can still have warrant for the proposition that I perceive something round. Again, even if I can't perceive colors at all, I can still have visual warrant for the proposition that something is red; even if for me nothing appears redly (everything is merely black and white) I might still be able to see that something is red, in the way in which one can see, on a black and white television, which boxer is wearing the red trunks. And of course there will be many more qualifications of this sort necessary:¹¹ suppose my belief is based upon two different mechanisms and one but not the other is functioning properly; suppose the same process works properly over one part of its range of operation but not over another, and my belief is produced by its working over both of these parts of its range of operation; or suppose a process is not working properly over part of its range but produces in me in given circumstances the very same belief it would have if it were working properly; in these cases does my belief have warrant? These are good questions, but there isn't time to work out all the answers here.

Still further, proper functioning, of course, *comes in degrees*; or if it does not, then approximation to proper functioning does. Clearly the faculties relevant with respect to a given belief need not be functioning *perfectly* for me to have warrant for my belief; many of my visual beliefs may constitute

knowledge even if my vision is not 20/20. Similarly, my faculties can function *properly* even if they do not function *ideally*, even if they do not function as well as those of some other actual or possible species (a point I discuss in chapter 6 of *Warrant: The Current Debate*). My locomotory equipment may be functioning properly even if I can't run as fast as a cheetah; my arithmetic powers may be in good working order even if I can't anywhere nearly keep up with a computer, or an angel, or an Alpha Centaurian. But how well, then, must such powers be functioning? Part of the answer here, of course, is that there is no answer; the ideas of knowledge and warrant are to some degree vague; hence there needs to be no precise answer to the question in question. What I hope is that the vagueness involved in my account of warrant vary with the vaguenesses we independently recognize in the notion of warrant. If warrant and proper function are properly tied together, then we may expect that they will waver together.

Similar comments and qualifications, of course, must be made about the environmental condition. For my beliefs to have warrant, the environment must be similar to that for which my epistemic powers have been designed; but just how similar must it be? Here, of course, we encounter vagueness; there is no precise answer. Further, suppose I *know* that the environment is misleading; and suppose I know in just which ways it is misleading. (I'm on a planet where things that look square are really round.) Then, clearly enough, the fact that my environment is misleading need not deprive my beliefs of warrant. And of course the same must be said for the requirement that my faculties be in good working order. Suppose (as in Castañeda's fantasy)¹² I suffer from a quirk of memory: whenever I read a history book, I always misremember the dates, somehow adding ten years to the date as stated: beliefs formed by way of reading history books – even beliefs about dates – can still have warrant for me; I can compensate for my erroneous tendency. What counts, of course, are uncorrected and uncompensated malfunctionings. Clearly there is need here for a good deal of Chisholming; let me postpone it, however, in order to turn to other more pressing matters.

II The Design Plan

But aren't there cases in which our faculties function perfectly properly in the right sort of

environment but the resulting beliefs still lack warrant? Surely there are. Someone may remember a painful experience as less painful than it was, as is sometimes said to be the case with childbirth.¹³ You may continue to believe in your friend's honesty long after evidence and cool, objective judgment would have dictated a reluctant change of mind. I may believe that I will recover from a dread disease much more strongly than is justified by the statistics of which I am aware. William James's climber in the Alps, faced with a life or death situation, believed more strongly than the evidence warrants that he could leap the crevasse. In all of these cases, there is no cognitive dysfunction or failure to function properly; it would be a mistake, however, to say that the beliefs in question had warrant for the person in question.

I cannot forbear quoting a couple of Locke's examples:

Would it not be an insufferable thing for a learned professor, and that which his scarlet would blush at, to have his authority of forty years standing wrought out of hard rock Greek and Latin, with no small expence of time and candle, and confirmed by general tradition, and a reverent beard, in an instant overturned by an upstart novelist? Can any one expect that he should be made to confess, that what he taught his scholars thirty years ago, was all error and mistake; and that he sold them hard words and ignorance at a very dear rate?¹⁴

The professor's faculties may be functioning properly (there may be a properly functioning defense mechanism at work); but his belief that the young upstart is dead wrong would have little by way of warrant. Another of Locke's examples:

Tell a man, passionately in love, that he is jilted; bring a score of witnesses of the falsehood of his mistress, 'tis ten to one but three kind words of hers, shall invalidate all their testimonies. . . . What suits our wishes, is forwardly believed is, I suppose, what every one hath more than once experimented; and though men cannot always openly gain-say, or resist the force of manifest probabilities, that make against them; yet yield they not to the argument. (*Essay*, IV, xx, 12)

Now it was widely believed in the eighteenth century that love was or induced a sort of madness, so

that the lover's epistemic faculties are not functioning properly. Even if that isn't so, however, even if we are designed to act and believe in extravagant fashion when in love, the lover's belief that his mistress is true to him has little by way of warrant.

Still another case: according to Freud, religious belief is "the universal obsessional neurosis of mankind"; religious belief consists in "illusions, fulfillments of the oldest, strongest, and most insistent wishes of mankind."¹⁵ Rather similar sentiments are expressed by Marx, who holds that religious belief is produced by an unhealthy, perverted social order: "This State, this society, produce religion, produce a perverted world consciousness, because they are a perverted world. . . . Religion is the sigh of the oppressed creature, the feelings of a heartless world, just as it is the spirit of unspiritual conditions."¹⁶ Now neither Freud nor Marx would be mollified if we pointed out that religion is very widespread among human beings, that is, "normal" in the statistical sense; what is statistically normal may still be a disease, a matter of malfunction, in this case a cognitive dysfunction. But there is a further subtlety here; Freud and Marx differ in a significant way. Marx seems to think that religion is a sort of perversion, something unhealthy; it is as if he says, "Let's call it an aberration and be done with it." Freud, on the other hand, is ambivalent. First, he says that religious belief is or stems from neurosis: that sounds like he thinks religious belief arises from a cognitive malfunction of some sort. But then he also says it is a matter of illusion, and arises from the "oldest and strongest and most insistent wishes of mankind." That suggests not that religious belief arises from malfunction or failure of some cognitive module to function properly, but instead by way of wish fulfillment. What one believes in *that* way isn't necessarily a product of malfunction; illusion and wish fulfillment also have their functions. According to Freud, they enable us to mask the grim, threatening, frightening visage of the world – a visage that would otherwise cause us to cower in terror or sink into utter and apathetic despair. On the second way of thinking, then, religious belief need not be a result of malfunction; it might be produced by faculties functioning just as they should. Even so, however – even if the wish fulfillment that produces religious belief does not result from cognitive malfunction – religious belief won't enjoy much by way of warrant.

So the proposed condition for warrant – proper function in an appropriate environment – isn't anywhere nearly sufficient for warrant. Why not? Well, consider the elements of our cognitive faculties responsible for beliefs of the above sorts – those produced by wishful thinking, or by the optimism that enables one to survive a deadly illness – one thinks that the purpose of *these* modules of our cognitive capacities is not to produce true beliefs. They are instead aimed at something else: survival, or the possibility of friendship, or (Freud thinks) the capacity to carry on in this bleak and nasty world of ours.

To get a better understanding of this matter, we must consider a notion of crucial importance: that of specifications, or blueprint, or *design plan*. Human beings are constructed according to a certain design plan. This terminology does not commit us to supposing that human beings have been literally designed – by God, for example. Here I use "design" the way Daniel Dennett (not ordinarily thought unsound on theism) does in speaking of a given organism as possessing a certain design, and of evolution as producing optimal design: "In the end, we want to be able to explain the intelligence of man, or beast, in terms of his design; and this in turn in terms of the natural selection of this design."¹⁷ We take it that when the organs (or organic systems) of a human being (or other organism) function properly, they function *in a particular way*. Such organs have a *function or purpose*; more exactly, they have several functions or purposes, including both proximate and more remote purposes. The ultimate purpose of the heart is to contribute to the health and proper function of the entire organism (some might say instead that it is to contribute to the *survival* of the individual, or the species, or even to the perpetuation of the genetic material itself).¹⁸ But of course the heart also has a much more circumscribed and specific function: to pump blood. Such an organ, furthermore, normally functions in such a way as to fulfill its purpose; but it also functions to fulfill that purpose in just one of an indefinitely large number of possible ways. Here a comparison with artifacts is useful. A house is designed to produce shelter – but not in just any old way. There will be plans specifying the length and pitch of the rafters, what kind of shingles are to be applied, the kind and quantity of insulation to be used, and the like. Something similar holds in the case of us and our faculties; we seem to be constructed in accordance with a specific set of plans. Better (since this

analogy is insufficiently dynamic) we seem to have been constructed in accordance with a set of specifications, in the way in which there are specifications for, for example, the 1992 Buick. According to these specifications (I'm just guessing), after a cold start the engine runs at 1,500 RPM until the engine temperature reaches 190°F; it then throttles back to 750 RPM.

Similarly, there is something like a set of specifications for a well-formed, properly functioning human being – an extraordinarily complicated and highly articulated set of specifications, as any first-year medical student could tell you. *Something* like such a set: a copy of these specifications does not come with every newborn child, and we can't write to the manufacturer for a new copy to replace the one we have carelessly lost. Suppose we call these specifications a "design plan." It is natural to speak of organisms and their parts as exhibiting design, and such talk is exceedingly common: "According to Dr Sam Ridgway, physiologist with the US Naval Ocean Systems Center in San Diego, seals avoid the bends by not absorbing nitrogen in the first place. 'The lungs of marine mammals,' Dr Ridgway explains, 'are designed to collapse under pressure exerted on deep dives. Air from the collapsed lungs is forced back into the windpipe, where the nitrogen simply can't be absorbed by the blood.'"¹⁹ Of course the design plan for human beings will include specifications for our *cognitive* system or faculties. Like the rest of our organs and systems, our cognitive faculties can work well or badly; they can malfunction or function properly. They too work in a certain way when they are functioning properly – and work in a certain way to accomplish their purpose. The purpose of the heart is to pump blood; that of our cognitive faculties (overall) is to supply us with reliable information: about our environment, about the past, about the thoughts and feeling of others, and so on. But not just any old way of accomplishing this purpose in the case of a specific cognitive process is in accordance with our design plan. It is for this reason that it is possible for a belief to be produced by a cognitive process or belief-producing mechanism that is *accidentally* reliable (as in the case of the processes I have cited as counterexamples to Goldman's version of reliabilism).²⁰ Although such belief-producing processes are in fact reliable, the beliefs they yield have little by way of warrant; and the reason is that these processes are pathologically out of accord with the design plan for human beings.

Our design plan, of course, is such that our faculties are highly responsive to circumstances. Upon considering an instance of *modus ponens*, I find myself believing its corresponding conditional; upon being appeared to in the familiar way, I find myself with the belief that there is a large tree before me; upon being asked what I had for breakfast, I reflect for a moment, and the belief that what I had was eggs on toast is formed within me. In these and other cases I do not *deliberate*; I do not total up the evidence (I am being appeared to redly; on most occasions when thus appeared to I am in the presence of something red; so most probably in this case I am) and thus come to a view as to what seems best supported; I simply find myself with the appropriate belief. Of course in *some* cases I may go through such a weighing of the evidence; for example, I may be trying to evaluate the alleged evidence in favor of the theory that human life evolved by means of such mechanisms as random genetic mutation and natural selection from unicellular life (which itself arose by substantially similar mechanical processes from nonliving material); but in the typical case of belief formation nothing like this is involved.

Here I wish to note just a couple of its salient features. According to our design plan, obviously enough, *experience* plays a crucial role in belief formation. *A priori* beliefs, for example, are not, as this denomination mistakenly suggests, formed prior to or in the absence of experience. Thinking of the corresponding conditional of *modus ponens* somehow *feels* different from thinking of, say, the corresponding conditional of *affirming the consequent*; and this difference in experience is connected with our accepting the one and rejecting the other. Of course experience plays a different role here from the role it plays in the formation of perceptual beliefs; it plays a still different role in the formation of memory beliefs, moral beliefs, beliefs about the mental lives of other persons, beliefs we form on the basis of inductive evidence, and the like. In later chapters we shall look into these matters in more detail.

Further, our design plan is such that under certain conditions we form one belief *on the evidential basis* of others. I may form the belief that Sam was at the party on the evidential basis of other beliefs – perhaps I learn from you that Sam wasn't at the bar and from his wife that he was either at the bar or at the party. Of course (if our faculties are functioning properly) we don't form

just *any* belief on the evidential basis of just any other. I won't form the belief that Feike is a Catholic on the evidential basis of the propositions that nine out of ten Frisians are Protestants and Feike is a Frisian – not, at any rate, unless I am suffering from some sort of cognitive malfunction. And here too experience plays an important role. The belief about Sam *feels like* the right one; that belief about Feike (in those circumstances) feels strange, inappropriate, worthy of rejection, not to be credited. Still further, the design plan dictates the appropriate *degree* or firmness of a given belief in given circumstances. You read in a relatively unreliable newspaper an account of a 53-car accident on a Los Angeles freeway; perhaps you then form the belief that there was a 53-car accident on the freeway. But if you hold that belief as firmly as, for example, that $2 + 1 = 3$, then your faculties are not functioning as they ought to and the belief has little warrant for you. Again, experience obviously plays an important role. What we need is a full and appropriately subtle and sensitive description of the role of experience in the formation and maintenance of all these various types of beliefs. For the moment, we may rest satisfied simply to note the importance of experience in the economy of our cognitive establishment.

Now return to the examples that precipitated this excursus about the design plan – the cases of beliefs produced by wish fulfillment, or the optimism necessary to surviving a serious illness, or willingness to have more children, or the like. In these cases, the relevant faculties may be functioning properly, functioning just as they ought to, but nevertheless not in a way that leads to truth, to the formation of true beliefs. But then proper function in a right environment is not sufficient for warrant. Different parts or aspects of our cognitive apparatus have different purposes; different parts or aspects of our design plan are aimed at different ends or goals. Not all aspects of the design of our cognitive faculties need be aimed at the production of true belief; some might be such as to conduce to survival, or relief from suffering, or the possibility of loyalty, or inclination to have more children, and so on. What confers warrant is one's cognitive faculties working properly, or working according to the design plan *insofar as that segment of the design plan is aimed at producing true beliefs*. But someone whose holding a certain belief is a result of an aspect of our cognitive design that is aimed not at truth but at something else won't be such that the belief has warrant for him; he won't prop-

erly be said to know the proposition in question, even if it turns out to be true.

So there are cases where belief-producing faculties are functioning properly but warrant is absent: cases where the design plan is not aimed at the production of true (or verisimilitudinous) beliefs but at the production of beliefs with some other virtue. But then there will also be cases where cognitive faculties are not functioning properly, but warrant is present; these will be inverses, so to speak, of the cases of the preceding paragraph. Suppose our design demands that under certain special circumstances our ordinary belief-producing mechanisms are overridden by a mechanism designed to deal with that specific case: perhaps there is a sort of optimistic mechanism that cuts in when I am seriously ill, causing me to believe more strongly than the evidence indicates that I will survive the illness, thereby enhancing my chances to survive it. Suppose I am taken seriously ill, and suppose through some malfunction (induced, perhaps, by the illness itself) the operation of the optimistic mechanism is inhibited, so that, believing just in accord with the evidence, I form the belief that I probably will not survive. Then the relevant segment of my cognitive faculties is not functioning properly; that is, it is not functioning in accordance with the design plan; but doesn't my belief have warrant anyway?²¹ Might I not have the degree of warrant that goes with the degree to which I believe that I probably won't survive, despite the fact that if my faculties were functioning properly, I would believe (to one or another degree of firmness) that I *will* survive? The answer, of course, is as before: those segments of my cognitive faculties (those modules, we might say) that are aimed at *truth* are functioning properly; my cognitive faculties are functioning in accord with the design plan insofar as the design plan is aimed at the production of true beliefs. There is malfunction only with respect to those cognitive modules aimed at something other than truth; so in this case the belief that I will not survive has the degree of warrant normally going with the degree of belief I display.

Many questions remain,²² but I must leave them to the reader.

III Reliability

According to the zeroth approximation, a belief has warrant for me, speaking roughly, if it is pro-

duced by my cognitive faculties functioning properly in a congenial environment. We have just seen that these two together are insufficient: the segment of the design plan governing the production of the belief in question must also be aimed at truth. But this is still insufficient. For suppose a well-meaning but incompetent angel – one of Hume’s infant deities,²³ say – sets out to design a variety of rational persons, persons capable of thought, belief, and knowledge. As it turns out, the design is a real failure; the resulting beings hold beliefs, all right, but most of them are absurdly false.²⁴ Here all three of our conditions are met: the beliefs of these beings are formed by their cognitive faculties functioning properly in the cognitive environment for which they were designed, and furthermore the relevant modules of the design plan are aimed at truth (the relevant modules of their cognitive equipment have the production of true beliefs as their purpose). But the beliefs of these pitifully deceived beings do not have warrant.²⁵ What must we add? That the design plan is a *good* one – more exactly, that the design governing the production of the belief in question is a good one; still more exactly, that the objective probability of a belief’s being true, given that it is produced by cognitive faculties functioning in accord with the relevant module of the design plan, is high. Even more exactly, the module of the design plan governing its production must be such that it is objectively highly probable that a belief produced by cognitive faculties functioning properly according to that module (in a congenial environment) will be true or verisimilitudinous. This is the reliabilist constraint on warrant, and the important truth contained in reliabilist accounts of warrant.

It is easy to overlook this condition. The reason is that we ordinarily take it for granted that when our cognitive faculties – at any rate, those whose function it is to produce true beliefs – function properly in an appropriate environment, then for the most part the beliefs they produce are true. When our faculties function in accord with our design plan (in an appropriate environment), the beliefs they produce are for the most part true. Certainly we think so with respect to memory, perception, logical and arithmetical beliefs, inductively based beliefs, and so on. Further, we take it for granted that these faculties are *reliable*; they not only *do* produce true beliefs, but *would* produce true beliefs even if things were moderately different. (They produce true beliefs in most of the

appropriately nearby possible worlds; that is, most of the appropriately nearby possible worlds *W* meet the following condition: necessarily, if *W* had been actual, then our cognitive faculties would have produced mostly true beliefs.) Still another way to put it: we take it for granted that the statistical or objective probability of a belief’s being true, given that it has been produced by our faculties functioning properly in the cognitive environment for which they were designed, is high. Perhaps more specifically our presupposition is that in general (for a person *S* with properly functioning faculties in an appropriate environment, and given the cited qualifications) the more firmly *S* believes *p*, the more likely it is that *p* is true. Of course, we think some faculties more reliable than others, and think a given faculty is more reliable under some conditions than others. This assumption on our part is a sort of presumption of reliability. Of course, it *is* a presumption or an assumption; it isn’t or isn’t obviously²⁶ entailed by the notion of proper function itself. So the account of proper function must include it as another condition: if one of my beliefs has warrant, then the module of the design plan governing the production of that belief must be such that the statistical or objective probability of a belief’s being true, given that it has been produced in accord with that module in a congenial cognitive environment, is high.

How high, precisely? Here we encounter vagueness again; there is no precise answer. It is part of the presumption, however, that the degree of reliability varies as a function of degree of belief. The things we are most sure of – simple logical and arithmetical truths, such beliefs as that I now have a mild ache in my knee (that indeed I have knees), obvious perceptual truths – these are the sorts of beliefs we hold most firmly, perhaps with the maximum degree of firmness, and the ones such that we associate a very high degree of reliability with the modules of the design plan governing their production. Even here, however, we are not immune from error: even what seems to be self-evident can be mistaken, as Frege learned to his sorrow.²⁷ It may be worth noting, however, that Frege did not believe the offending “axiom” to the maximal degree; if he had, then he would have been no more likely to give up that “axiom” than to conclude that there really *is* a set that is and is not a member of itself.

I say the presupposition of reliability is a feature of our usual way of thinking about warrant; but of

course this presupposition is not inevitable for us. The skeptic, for example, can often best be seen as questioning this presupposition. She may agree that there is indeed a perfectly proper distinction between cognitive proper function and malfunction, but be agnostic about the question whether there is any correlation at all between proper function and truth. Or she may think there is indeed such a correlation, but think it far too weak to support our ordinary claims to knowledge. Or she may think that since the long-run purpose of our beliefs, as she sees it, is to enable us to move about in the environment in such a way that we do not come to grief (or do not come to grief until we have had a chance to reproduce), there is no interesting correlation between a belief's being produced by faculties functioning properly and its being true.²⁸ Of course one can be a skeptic about one particular area as opposed to others: a rationalist may think sense perception less reliable than reason and may thus maintain that it is only reason, not perception, that gives us knowledge; an empiricist may see things the other way around. Philosophy itself is a good candidate for a certain measured skepticism: in view of the enormous diversity of competing philosophical views, one can hardly claim with a straight face that what we have in philosophy is *knowledge*; the diversity of views makes it unlikely that the relevant segments of the design plan are sufficiently reliable. (In a

properly run intellectual establishment, therefore, most philosophical views will not enjoy anywhere nearly the maximal degree of belief.)

To return to warrant then: to a first approximation, we may say that a belief *B* has warrant for *S* if and only if the relevant segments (the segments involved in the production of *B*) are functioning properly in a cognitive environment sufficiently similar to that for which *S*'s faculties are designed; and the modules of the design plan governing the production of *B* are (1) aimed at truth, and (2) such that there is a high objective probability that a belief formed in accordance with those modules (in that sort of cognitive environment) is true; and the more firmly *S* believes *B* the more warrant *B* has for *S*. This is at best a first approximation; it is still at most programmatic, a suggestion, an idea, a hint. Furthermore, it might be suggested (in fact, it *has* been suggested) that while it may be difficult to find counterexamples to the view, that is only because it is vague and imprecise. I have sympathies with both complaints, although I would implore those who make the second to heed Aristotle's dictum and seek no more precision than the subject admits. Maybe there isn't any neat formula, any short and snappy list of conditions (at once informative and precise) that are severally necessary and jointly sufficient for warrant; if so, we won't make much progress by grimly pursuing them.

Notes

- 1 "Evidentialism," this vol., ch. 16.
- 2 See his "Concepts of Epistemic Justification," *Monist* (January 1985), and "An Internalist Externalism," *Synthese* 74, no. 3 (1988); see also several of the articles collected in *Epistemic Justification* (Ithaca: Cornell University Press, 1989).
- 3 Fred Dretske, *Explaining Behavior* (Cambridge: MIT Press, 1988), p. 91.
- 4 "Limiting Science: a Biologist's Perspective," *Daedalus* (Summer 1988), p. 336.
- 5 *The Child's Conception of Physical Causality* (London: Kegan Paul, Trench, Trubner, 1930).
- 6 "How to Build a Person," in *Philosophical Perspectives, 1, Metaphysics, 1987*, ed. James Tomberlin (Atascadero, CA.: Ridgeview, 1987), p. 146.
- 7 See my *Warrant: The Current Debate* (New York: Oxford University Press, 1993), ch. 6, sec. 1, "The Varieties of Rationality."
- 8 See *ibid.*, ch. 6, pp. 117ff.

- 9 It is sometimes suggested that whenever I believe *A* no more firmly than not-*A* and not-*A* no more firmly than *A*, then I can be thought of as believing *A* (and not-*A*) to degree .5. This seems clearly mistaken. Consider a case where I have no idea at all whether the proposition in question is true. You ask me (a touch pedantically) "consider the proposition that the highest mountain on Mars is between ten and eleven thousand feet high and call it 'A'; do you think *A* is true?" I have no idea about *A* and do not believe it more likely than its denial; I also do not believe its denial more likely than it. Then on the Bayesian view, I must believe *A* to degree .5. You then ask me the same question about *B*: the proposition that the highest mountain on Mars is between eleven and twelve thousand feet high. Again, I have no idea; so on the Bayesian view I am considering, I must also believe *B* to degree .5. Now *A* and *B* are mutually exclusive; according to the probability calculus, therefore, I should believe their disjunction to degree 1. But of

- course I do not; for I also have no idea whether the highest mountain on Mars is between ten and twelve thousand feet high. And the problem is not that I am desperately incoherent. The problem is that we can't properly represent ignorance of this sort as believing the proposition in question to degree .5. There is a vast difference between the situation in which I think *A* probable to degree .5 (perhaps *A* is the proposition that the die will come up side 1, 2, or 3) and the situation in which I have no idea what the probability of *A*'s being true might be.
- 10 See H. N. Castañeda's "The Indexical Theory of Knowledge," in Peter French, Theodore E. Uehling, Jr., and Howard Wettstein (eds), *Midwest Studies in Philosophy*, vol. V (Minneapolis: University of Minnesota Press, 1980).
 - 11 Here I am indebted to Tom Senor.
 - 12 "The Indexical Theory of Knowledge," p. 202.
 - 13 "A woman giving birth to a child has pain because her time has come; but when her baby is born she forgets the anguish because of her joy that a child is born into the world." John 16:21.
 - 14 *An Essay concerning Human Understanding*, ed. A. C. Fraser (New York: Dover, 1953), IV, xx, 11, hereafter referred to as *Essay*.
 - 15 *The Future of an Illusion* (1927), trans. and ed. James Strachey (London: Norton, 1961), p. 30.
 - 16 K. Marx, *Introduction to a Critique of the Hegelian Philosophy of Right*, in *Collected Works*, by K. Marx and F. Engels (London: Lawrence & Wishart, 1975), 3:175.
 - 17 *Brainstorms* (Cambridge: Bradford Books, 1978), p. 12.
 - 18 See Richard Dawkins, *The Selfish Gene* (Oxford: Oxford University Press, 1976).
 - 19 *National Geographic* 171, no. 4 (April 1987), p. 489.
 - 20 See my *Warrant: The Current Debate*, ch. 9.
 - 21 I owe this example to Caleb Miller.
 - 22 Some of which were forcibly brought to my attention by Dean Zimmerman.
 - 23 *Dialogues Concerning Natural Religion*, Part V (1779), ed. N. K. Smith (Bobbs-Merrill, 1947), p. 169.
 - 24 Some (Donald Davidson, for example) apparently hold that it is impossible that there be a sizeable community of believers most of whose beliefs are false; I disagree and explain why elsewhere.
 - 25 This counterexample was called to my attention by Richard Swinburne, Ian Foster, and Thomas Senor.
 - 26 I suppose it might sensibly be held that it is impossible that there be rational beings (beings capable of reasoning or belief) whose cognitive faculties function properly but who nonetheless hold predominantly false beliefs. Perhaps there are purposes or ends necessarily built into certain kinds of creatures. Then if a malevolent Cartesian demon were to design a race of rational creatures whose beliefs were nearly always mistaken, their cognitive faculties would not be functioning properly, even if they were functioning just as they were designed to. Instead, we should have to say that what this demon wanted to do was to design a race of cognitive beings that did not function properly.
 - 27 Frege produced a set of axioms for set theory, including the famous or infamous proposition that for any property *P* there exists the set of just those things that have *P*. Russell showed him that this axiom (together with the others) yields a contradiction: if it is true, there will be a set of nonself-membered sets, which both will and will not be a member of itself.
 - 28 Thus Patricia Churchland: "Boiled down to essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing, fighting and reproducing. The principal chore of nervous systems is to get the body parts where they should be in order that the organism may survive. . . . Truth, whatever that is, definitely takes the hindmost" (*Journal of Philosophy* 84 (October 1987), p. 548).

Virtues of the Mind

Linda Zagzebski

General Account of a Virtue

A serious problem in any attempt to give a general account of the nature of virtue is that our language does not contain a sufficient number of names that convey the full unified reality of each virtue. Some names pick out reactive feelings (empathy), some pick out desires (curiosity), some pick out motivations to act (benevolence), whereas others pick out patterns of acting that appear to be independent of feeling and motive (fairness). For this reason it is easy to confuse a virtue with a feeling in some cases (empathy, compassion), and with a skill in others (fairness). The result is that it is very difficult to give a unitary account of virtues using common virtue language. MacIntyre (1984) blames the problem on a defect in our culture,¹ but this cannot be an adequate explanation since Aristotle's list was no better in this respect than ours. When we examine Aristotle's virtues and vices we see that he had difficulty in finding names for some of them, and a few of his names seem forced, such as his term "*anesthesia*," which he coins for the trait of insensibility to pleasure. Gregory Trianosky's response to this situation is to say that virtues are not all traits of the same general type.² Robert Roberts also concludes that there are several distinct kinds of virtue.³ This response is understandable and it is possible that we will eventually be forced into it, but I believe it should only be taken as a last resort, and I see no reason to take it yet. It is more plausible that the problem derives from a

defect in our virtue language rather than a division in the nature of virtue itself.

Let us begin by reviewing the features of virtue we have already identified. First, a virtue is an acquired excellence of the soul, or to use more modern terminology, it is an acquired excellence of the person in a deep and lasting sense. A vice is the contrary quality; it is an acquired defect of the soul. One way to express the depth required for a trait to be a virtue or a vice is to think of it as a quality we would ascribe to a person if asked to describe her after her death. Perhaps no quality is really permanent, or, at least, no interesting quality, but virtues and vices are in the category of the more enduring of a person's qualities, and they come closer to defining who the person is than any other category of qualities.

Second, a virtue is acquired by a process that involves a certain amount of time and work on the part of the agent. This is not to suggest that a person controls the acquisition of a virtue entirely; that is plainly false. Nevertheless, the time and effort required partly account for a virtue's deep and lasting quality, one that in part defines a person's identity and that leads us to think of her as responsible for it. This means that typically a virtue is acquired through a process of habituation, although the virtues of creativity may be an exception.

Third, a virtue is not simply a skill. Skills have many of the same features as virtues in their manner of acquisition and in their area of application, and virtuous persons are expected to have the correlative skills in order to be effective in action, but skills do not have the intrinsic value of virtues.

Originally published in L. Zagzebski, *Virtues of the Mind* (Cambridge: Cambridge University Press, 1996), pp. 134-7, 166-84.

Fourth, a virtue has a component of motivation. A motivation is a disposition to have a certain motive, and a motive is an emotion that initiates and directs action to produce an end with certain desired features. Motivations can become deep parts of a person's character and provide her with a set of orientations toward the world that emerge into action given the appropriate circumstances. A motivation is best defined, not as a way of acting in circumstances specifiable in advance, but in terms of the end at which it aims and the emotion that underlies it. The easiest way to identify a motivation is by reference to the end at which it aims, but it also involves an emotion disposition, and that is harder to identify by name.

This brings us to another important feature of virtue: "Virtue" is a success term. The motivational component of a virtue means that it has an end, whether internal or external. A person does not have a virtue unless she is reliable at bringing about the end that is the aim of the motivational component of the virtue. For example, a fair person acts in a way that successfully produces a state of affairs that has the features fair persons desire. A kind, compassionate, generous, courageous, or just person aims at making the world a certain way, and reliable success in making it that way is a condition for having the virtue in question. For this reason virtue requires knowledge, or at least awareness, of certain nonmoral facts about the world. The nature of morality involves, not only wanting certain things, but being reliable agents for bringing those things about. The understanding that a virtue involves is necessary for success in bringing about the aim of its motivational component. This means that virtue involves a component of understanding that is implied by the success component.

A virtue therefore has two main elements: a motivational element, and an element of reliable success in bringing about the end (internal or external) of the motivational element. These elements express the two distinct aims of the moral project that we find in commonsense moral thinking. On the one hand, ordinary ways of thinking about morality tell us that morality is largely a matter of the heart, and we evaluate persons for the quality of their motivations. But morality is also in part a project of making the world a certain kind of place – a better place, we might say, or the kind of place good people want it to be. Because of the latter interest, we are impressed with moral success, not to the exclusion of an

interest in people's cares and efforts, but in addition to it.

A virtue, then, can be defined as a deep and enduring acquired excellence of a person, involving a characteristic motivation to produce a certain desired end and reliable success in bringing about that end. What I mean by a motivation is a disposition to have a motive; a motive is an action-guiding emotion with a certain end, either internal or external.

This definition is broad enough to include the intellectual as well as the traditional moral virtues. It may also be broad enough to include virtues other than the moral or intellectual, such as aesthetic, religious, or perhaps even physical virtues, but I will not consider virtues in these other categories in this work. The definition may not apply to higher-order virtues such as integrity and practical wisdom, however.

The Motivation for Knowledge and Reliable Success

In this section I will argue that the individual intellectual virtues can be defined in terms of motivations arising from the general motivation for knowledge and reliability in attaining the aims of these motives. Since all of the intellectual virtues have the same foundational motivation and since all of the other moral virtues have different foundational motivations, this means that a distinction between an intellectual and a moral virtue can be made on the basis of the motivational component of the virtue. I maintain that this is the only theoretically relevant difference between intellectual virtues and the other moral virtues, and so there are good grounds for continuing to call these virtues "intellectual," even though I have argued that they are best treated as a subset of the moral virtues. It may be that at the deepest level the moral and intellectual virtues arise from the same motivation, perhaps a love of being in general.⁴ If so, such a motivation would serve to unify all the virtues, but I will not analyze the relations among the virtuous motivations in this work.

The simplest way to describe the motivational basis of the intellectual virtues is to say that they are all based in the motivation for knowledge. They are all forms of the motivation to have cognitive contact with reality, where this includes more than what is usually expressed by saying that people desire truth. Understanding is also a

form of cognitive contact with reality, one that has been considered a component of the knowing state in some periods of philosophical history. I will not give an account of understanding in this work, but I have already indicated that it is a state that includes the comprehension of abstract structures of reality apart from the propositional. I will assume that it either is a form of knowledge or enhances the quality of knowledge. Although all intellectual virtues have a motivational component that aims at cognitive contact with reality, some of them may aim more at understanding, or perhaps at other epistemic states that enhance the quality of the knowing state, such as certainty, than at the possession of truth per se. A few stellar virtues such as intellectual originality or inventiveness are related, not simply to the motivation for the *agent* to possess knowledge, but to the motivation to advance knowledge for the human race. We will also look at how the motivation to know leads to following rules and belief-forming procedures known by the epistemic community to be truth conducive, and we will see how the individual intellectual virtues are knowledge conducive.

The task of defining virtues immediately raises the question of how virtues are individuated and whether they are unified at some deeper level. I will not go very far into this matter, although it is an interesting one and ought to be pursued in a full theory of virtue. I have no position on the question of whether intellectual virtues that share a name with certain moral virtues are two different virtues or one. Even within the class of intellectual virtues it is difficult to demarcate the boundaries of the individual virtues if I am right that they all arise out of the motivation for knowledge since that implies that all intellectual virtues are unified by one general motivation. But, of course, the same thing can be said about all the other moral virtues since they also can be unified by one general motivation for good, and knowledge is a form of good.

Let me address one more point before beginning. The definition of intellectual virtue in terms of the motivation for knowledge is circular if we then go on to define knowledge in terms of intellectual virtue. The thesis here must be formulated less succinctly but without circularity as the thesis that the individual intellectual virtues can be defined in terms of derivatives of the motivations for truth or cognitive contact with reality, where the motivation for understanding is assumed to be a form of the motivation for cognitive contact with

reality. I am formulating the position in terms of the motivation for knowledge because I think that that is closer to the way people actually think of their own motives and the way those motives are described by others, but I am not wedded to this view. The formulation in terms of knowledge motivation is simpler, and, of course, it is only circular when the theory of virtue is combined with the theory of knowledge.

The motivation for knowledge

Intellectual virtues have been neglected in the history of philosophy, but there were discussions of them in the early modern period as part of the general critical examination of human perceptual and cognitive faculties that dominated that era. Both Hobbes and Spinoza connected the intellectual as well as the moral virtues with the passions, and both traced the source of these virtues to a single human motivation, the motivation for self-preservation or power. In the early part of this century John Dewey stressed the place of the intellectual virtues in what he called "reflective thinking," arising from the desire to attain the goals of effective interaction with the world. We will look first at some remarks by Hobbes and Dewey, and then I will turn to the contemporary treatment of the intellectual virtues by James Montmarquet in the course of giving my own argument for the derivation of the motivational components of intellectual virtues from the motivation to know.

Let us begin with the lively discussion of the causes of intellectual virtue and vice in Hobbes's *Leviathan*:

The causes of this difference of wits are in the passions, and the difference of passions proceeded partly from the different constitution of the body and partly from different education. For if the difference proceeds from the temper of the brain and the organs of sense, either exterior or interior, there would be no less difference of men in their sight, hearing, or other sense than in their fancies and discretions.⁵ It proceeds, therefore, from the passions, which are different not only from the difference of men's complexions, but also from their difference of customs and education.

The passions that most of all cause the difference of wit are principally the more or less

desire of power, of riches, of knowledge, and of honor. All which may be reduced to the first – that is, desire of power. For riches, knowledge and honor are but several sorts of power.

And therefore a man who has no great passion for any of these things but is, as men term it, indifferent, though he may be so far a good man as to be free from giving offense, yet he cannot possibly have either a great fancy or much judgment. For the thoughts are to the desires as scouts and spies, to range abroad and find the way to the things desired, all steadiness of the mind's motion, and all quickness of the same, proceeding from thence; for as to have no desire is to be dead, so to have weak passions is dullness; and to have passions indifferently for everything, GIDDINESS and *distraction*; and to have stronger and more vehement passions for anything than is ordinarily seen in others is that which men call MADNESS.⁶

A couple of points in this passage are of interest to our present concern. First, the motivation for knowledge is not a basic motive but is a form of the motivation for power, according to Hobbes. Second, Hobbes's cognitively ideal person is not passionless, but cognitive defects can be traced to an excessively strong, excessively weak, or misplaced desire for power. I will not question the first point. I think Hobbes is probably wrong in his reduction of the desire for knowledge to the desire for power, but I will not dispute it here since even if he is right, the effect is simply to add another motivational layer beneath the one I am proposing, and so it is no threat to the structure of the theory I am proposing. But I want to call attention to Hobbes's second point, which I find insightful. Hobbes says that cognitive virtues and vices arise from differences in a motivation, and that motivation is a passion that admits of excess, deficiency, and distortion of various sorts, and this seems to me to be generally right. I differ with Hobbes mainly in that I identify this motivation with the motivation for knowledge, whereas Hobbes includes several other forms of the motivation for power along with the motivation for knowledge.

If the human drive for knowledge naturally and inexorably led to success, there would be no need for intellectual virtues. But this motivation can be deficient or distorted in many ways, leading to intellectual vices. Deficiency is presumably one of the most common problems, and Ralph Waldo Emerson expresses a pessimistic view of the

human drive for knowledge that illustrates how a natural human motivation can be affected by lethargy:

God offers to every mind its choice between truth and repose. Take which you please, – you can never have both. Between these, as a pendulum, man oscillates. He in whom the love of repose predominates will accept the first creed, the first philosophy, the first political party he meets, – most likely his father's. He gets rest, commodity, and reputation; but he shuts the door to truth. He in whom the love of truth predominates will keep himself aloof from all moorings, and afloat. He will abstain from dogmatism, and recognize all the opposite negations between which, as walls, his being is swung. He submits to the inconvenience of suspense and imperfect opinion, but he is a candidate for truth, as the other is not, and respects the highest law of his being. (“Intellect,” Essay 11)

In this passage Emerson describes how a deficiency in the desire for truth leads to such cognitive vices as lack of autonomy, closed-mindedness, and dogmatism. This may lead us to wonder whether an excess of the motivation for knowledge can also lead to intellectual vices, as Hobbes implies in the passage quoted above. This is parallel to the question of whether a person can be a moral fanatic: excessively motivated by a desire to do or to produce good. Since it is problematic whether this is possible, we will not examine it here.

Few philosophers have given positive directions on how to think that are intended to circumvent the pitfalls in forming beliefs. The stress has generally been on the mistakes. A well-known exception is Descartes in *Rules for the Direction of the Mind*, and another is John Dewey in *How We Think*. I will not discuss the former since it has been exhaustively examined many times, but I find Dewey intriguing if rather nonspecific. Although he does not discuss the motivation for knowledge directly, he does discuss the motivations to reach our goals in action and to make systematic preparations for the future and the desire to be free from the control of nature, all of which are closely connected with knowledge.⁷ These values require the practice of what Dewey calls “reflective thinking,” which he outlines in some detail:

No one can tell another person in any definite way how he *should* think, any more than how he ought to breathe or to have his blood circulate. But the various ways in which we *do* think can be told and can be described in their general features. Some of these ways are better than others; the reasons why they are better can be set forth. The person who understands what the better ways of thinking are and why they are better can, if he will, change his own personal ways until they become more effective; until, that is to say, they do better the work that thinking can do and that other mental operations cannot do so well. The better way of thinking that is to be considered in this book is called reflective thinking. (p. 3)

The disclaimer in the first sentence of the above passage is surely too strong, but the rest of the paragraph is reasonable. Dewey goes on to say that reflective thinking requires not only certain skills, but also certain "attitudes":

Because of the importance of attitudes, ability to train thought is not achieved merely by knowledge of the best forms of thought. Possession of this information is no guarantee for ability to think well. Moreover, there are no set exercises in correct thinking whose repeated performance will cause one to be a good thinker. The information and the exercises are both of value. But no individual realizes their value except as he is personally animated by certain dominant attitudes *in his own character* [emphasis added]. It was once almost universally believed that the mind had faculties, like memory and attention, that could be developed by repeated exercise, as gymnastic exercises are supposed to develop the muscles. This belief is now generally discredited in the large sense in which it was once held. . . .

What can be done, however, is to cultivate those *attitudes* that are favorable to the use of the best methods of inquiry and testing. Knowledge of the methods alone will not suffice; there must be the desire, the will, to employ them. This desire is an affair of personal disposition. But on the other hand the disposition alone will not suffice. There must also be understanding of the forms and techniques that are the channels through which these attitudes operate to the best advantage. (pp. 29–30)

In this passage Dewey places special importance on the desire to employ better ways of thinking, claiming that knowledge of methods is not sufficient. He thus traces a path from our motivation to believe truly and to act effectively to the formation of "attitudes" or intellectual virtues that lead us to employ certain methods of thinking and forming beliefs. For my purposes, the salient point is that the foundation of these virtues is a motivation: the motivation to think more effectively.

The "attitudes" Dewey says one needs to cultivate are the following:

Open-mindedness. "This attitude may be defined as freedom from prejudice, partisanship, and such other habits as close the mind and make it unwilling to consider new problems and entertain new ideas" (p. 30).

Wholeheartedness. "When a person is absorbed, the subject carries him on. Questions occur to him spontaneously; a flood of suggestions pour in on him; further inquiries and readings are indicated and followed; instead of having to use his energy to hold his mind to the subject. . . . the material holds and buoys his mind up and gives an onward impetus to thinking. A genuine enthusiasm is an attitude that operates as an intellectual force. A teacher who arouses such an enthusiasm in his pupils has done something that no amount of formalized method, no matter how correct, can accomplish" (pp. 31–2).

Responsibility. "Like sincerity or wholeheartedness, responsibility is usually conceived as a moral trait rather than as an intellectual resource. But it is an attitude that is necessary to win the adequate support of desire for new points of view and new ideas and of enthusiasm for and capacity for absorption in subject matter. These gifts may run wild, or at least they may lead the mind to spread out too far. They do not of themselves insure that centralization, that unity, which is essential to good thinking. To be intellectually responsible is to consider the consequences of a projected step; it means to be willing to adopt these consequences when they follow reasonably from any position already taken. Intellectual responsibility secures integrity; that is to say, consistency and harmony in belief" (p. 32).

In the contemporary literature Laurence Bonjour and Hilary Kornblith⁸ introduced a motivational element into the discussion of epistemic normativity in the notion of epistemic responsibility, defined by Kornblith as follows: "An *epistemically responsible agent* desires to have true beliefs, and thus desires to have his beliefs produced by processes which lead to true beliefs; his actions are guided by these desires" (p. 34). Although Kornblith does not specifically discuss intellectual virtues, he implies that a motivation or desire is at the root of the evaluation of epistemic agents, and that seems to me to be right. A more extensive treatment of epistemic virtue and its connection with motivation has been given by James Montmarquet⁹ who connects a large set of intellectual virtues with the desire for truth, claiming that these virtues are qualities a person who wants the truth would want to acquire. However, it is not Montmarquet's intention to define intellectual virtues the way I am proposing here or to derive them all from the motivation for truth or from the motivation for knowledge. Still, Montmarquet's work has an obvious affinity with the theory I am proposing. I want to give it close attention.

Recall Montmarquet's classification of the epistemic virtues. Briefly, they are the virtues of impartiality, or openness to the ideas of others; the virtues of intellectual sobriety, or the virtues of the careful inquirer who accepts only what is warranted by the evidence, and the virtues of intellectual courage, which include perseverance and determination. Notice that there is quite a bit of overlap between these sets of virtues and Dewey's. The major differences are in Dewey's virtue of wholeheartedness and Montmarquet's virtues of courage.

Montmarquet calls the desire for truth "epistemic conscientiousness" and argues that *some* intellectual virtues arise out of this desire.

The first point to be made . . . is that such qualities as open-mindedness are widely regarded as truth-conducive. In contrast to the highly controversial claims of various theories, the truth-conduciveness of qualities such as openness and intellectual sobriety is widely acknowledged to be a fact, not only by the expert (if there are "experts" on any such matter as this), but also by the average nonexpert individual (at least if he or she is suitably queried). Take openness. Unless one starts from the unlikely presumption

that one has found the truth already and that the contrary advice and indications of others is liable, therefore, only to lead one astray, one can hardly possess a sincere love of truth, but no concern about one's own openness. Or take intellectual sobriety. Here, too, unless one starts from the unlikely presumption that one's immediate reactions and unchecked inferences are so highly reliable as not to be improved by any tendency to withhold full assent until they are further investigated, the virtue of sobriety will have to be acknowledged. Or, finally, take intellectual courage. Again, unless one makes an initially unappealing assumption that one's own ideas – true as they may seem to oneself – are so liable to be mistaken as to require not only deference to the opinions of others, but also a deep sense that these are opinions more liable to be correct than one's own (even when one cannot see how or why) [, unless] one makes such an initial assumption, one will have to acknowledge intellectual courage as a virtue.¹⁰

The reader should not be misled into thinking that this is an argument that these virtues are truth conducive; in fact, Montmarquet questions the truth conduciveness of openness and courage, as we will see. It is, instead, an argument that they are traits persons who *desire* the truth would want to have. I take this to mean that such persons would be *motivated* to act the way open-minded, intellectually sober, cautious, courageous, and persevering people act in their belief-forming processes. So if a person is motivated to get the truth, she would be motivated to consider the ideas of others openly and fairly, to consider the evidence with care, not to back down too quickly when criticized, and all the rest. This seems to me to be correct. It means that the motivation for knowledge gives rise to the motivation to act in ways that are distinctive of the various intellectual virtues Montmarquet mentions. Undoubtedly it also leads to the motivation to acquire Dewey's trait of intellectual responsibility; in fact, the motivation to be able to accurately predict consequences is a form of the motivation to know. The trait that Dewey calls "wholeheartedness," the attitude of enthusiasm, which moves us onward in thinking, is also a form of the motivation to know, in fact, an intensification of it. It is reasonable to conclude, then, that a wide range of intellectual virtues arise out of the same *general* motivation, the motivation for knowledge, and have the same general aim, knowledge.

The success component of the intellectual virtues

Intellectually virtuous motivations lead the agent to guide her belief-forming processes in certain ways. They make her receptive to processes known to her epistemic community to be truth conducive and motivate her to use them, even if it means overcoming contrary inclinations. As Dewey tells us, it is not enough to be aware that a process is reliable; a person will not reliably *use* such a process without certain virtues. At least this is the case with reliable processes that are not unconscious or automatic. Contemporary research in epistemology has focused extensively on the concept of a truth-conducive belief-forming process, as well as on many specific examples of these processes. I have no intention of duplicating or replacing this work here. My purpose is to point out that the motivation for knowledge leads a person to follow rules and belief-forming processes that are truth conducive and whose truth conduciveness she is able to discover and use by the possession of intellectual virtue.

Intellectually virtuous motivations not only lead to following reliable procedures but also lead to the development of particular skills suited to the acquisition of knowledge in a certain area. Skills are more closely connected to effectiveness in a particular area of life or knowledge than are virtues, which are psychically prior and provide the motivations to develop skills. Intellectual skills are sets of truth-conducive procedures that are acquired through habitual practice and have application to a certain area of truth acquisition. Since the path to knowledge varies with the context, the subject matter, and the way a community makes a division of intellectual labor, people with the same intellectual virtues will not all need to have the same skills, at least not to the same degree. Clearly the importance of fact-finding skills, skills of spatial reasoning, and skills in the subtler branches of logic are not equally important for all areas of the pursuit of knowledge. But all of these skills could arise in different people from the same intellectual virtues – for example, carefulness, thoroughness, and autonomy.

We have already seen that virtue is more than a motivation. Of course, we would expect many virtuous motivations to lead to success in carrying out the aims of the motive. So, for example, the motive to be careful or persevering probably leads somewhat reliably to success in being careful or

persevering, but the correlation with success is probably much less in the case of such virtuous motives as the motive to be autonomous, the motive to be courageous, and perhaps even the motive to be open-minded. The weak connection between motive and success is also noticeable in Dewey's virtue of wholeheartedness (if it is a virtue), since it is surely naive to think that the motivation to be enthusiastic reliably leads to being enthusiastic. But even when the motivational component of a virtue is *generally* related to success, we do not call a person virtuous who is not reliably successful herself, whether or not most people who have the trait are successful in carrying out the aims of the virtue in question. So if she is truly open-minded, she must actually be receptive to new ideas, examining them in an evenhanded way and not ruling them out because they are not her own; merely being motivated to act in these ways is not sufficient. Similarly, if she is intellectually courageous, she must, in actual fact, refrain from operating from an assumption that the views of others are more likely to be true than her own and must be willing to withstand attack when she has good reason to think she is right, but not otherwise. Parallel remarks apply to the other intellectual virtues. It follows that each of these intellectual virtues has a motivational component arising out of the motivation to know and a component of reliable success in achieving the aim of the motivational component.

Most virtues are acquired by habituation and we only consider them virtues when they are entrenched in the agent's character. Entrenchment is a necessary feature of virtues because they are often needed the most when they encounter resistance. For example, the tendency to be motivated by compassion does not signify the existence of the *virtue* of compassion in a person who loses this motivation in the presence of physically unattractive persons in need, even if these circumstances do not arise very often. Similarly, the tendency to be motivated to fairly evaluate the arguments of others does not signify the existence of the virtue of intellectual fairness in a person who loses this motivation when confronted with arguments for unappealing conclusions, even if she is lucky enough not to encounter such arguments very often. So the motivational component of a virtue must be inculcated sufficiently to reliably withstand the influence of contrary motivations when those motivations do not themselves arise from virtues. The more that virtuous motivations and

the resulting behavior become fixed habits, the more they are able to reliably achieve the ends of the virtue in those cases in which there are contrary tendencies to be overcome.

One way to distinguish among the truth-conducive qualities those that are virtues and those that are not is by the difference in the value we place on the entrenchment of these traits. Montmarquet mentions that we would not want the desire to uphold behaviorist psychology to be an entrenched trait even if it is truth conducive,¹¹ unlike the desire for the truth itself or, I would add, the desire to be open-minded, careful in evaluating evidence, autonomous, etc. The latter traits, when entrenched, lead to the truth partly *because* of their entrenchment, whereas the desire to uphold behaviorism is less likely to lead to the truth if it is entrenched than if it is not. The intellectual virtues are a subset of truth-conducive traits that are entrenched and whose entrenchment aids their truth conduciveness.¹² The value of the entrenchment of a trait would, of course, depend partly on the environment in which it is entrenched.¹³ Most of the qualities I have been calling intellectual virtues – traits such as open-mindedness, carefulness, and perseverance – are to a great extent environment neutral, but this does not mean that there are not other intellectual virtues that are more context sensitive.

Many intellectual virtues, including those mentioned by Dewey, not only arise from and serve the motivation to know the truth, but are also crucial in such activities as the arts, crafts, and games. The ultimate aim of these activities is not knowledge but something practical: creating an artistically superior sonnet, making a fine violin, winning a chess game.¹⁴ These ends cannot be successfully achieved without knowledge in one of its senses, but probably not the kind of knowledge whose object is true propositions. At least, that sort of knowledge is not the one most fundamentally connected to success in these activities, which is more a matter of knowing-how rather than knowing-that. Still, some of the same virtues that arise out of the desire for knowledge and aid its successful achievement can also aid the achievement of these practical ends and, in some people, may arise more out of a desire for the practical end than out of a desire for knowledge. I do not claim, then, that intellectual virtues arise only from the motivation to know, much less do I claim they arise only from the motivation to have propositional knowledge, and I certainly do not claim that

their exercise is properly directed only at knowledge. The value of intellectual virtues extends beyond their epistemic use. So not only is the distinction between intellectual and moral virtues highly artificial, but the distinction between intellectual virtues and the practical virtues needed for doing such things as creating sonnets, making violins, or winning chess games is artificial as well. Again, I will not discuss the problem of virtue individuation. There may be *some* difference between, say, the kind of openness displayed in writing a Shakespeare sonnet and the kind of openness displayed in pure scientific investigation. This difference may amount to a distinction in the virtues themselves if virtue identity is determined by the ultimate end of the virtue. The point is that even if this is the case, there are practical and intellectual virtues so similar to each other that they are very difficult to distinguish, and this means that it is highly implausible to maintain that intellectual virtues are fundamentally different in kind from the virtues needed for the kinds of practical activities just named.

Amelie Rorty points out that while the utility and success of intellectual virtues depend on their becoming habits that lead to action without prior deliberation, habits can become pathological or idiotic.¹⁵ They become pathological, she says, when they become so habitual that their exercise extends to situations that no longer concern their internal aims. So generosity is pathological when it debilitates its recipients. The capacity to generate what Rorty calls “bravura virtuoso thought experiments” becomes pathological when it applies only to a very rare, narrow range of circumstances (p. 13). A virtuous habit becomes idiotic when its exercise resists a reasonable redirection of its aims, a redirection that is appropriate to changing circumstances. Rorty gives the example of courage when one is unable to make the transition from its military use to its use in political negotiation. In the intellectual sphere, the virtue of properly arguing from authority becomes idiotic when it is used to block the investigation of the legitimacy of the authority itself (p. 14). Some of these problems can be addressed by the function of the virtue of *phronesis*, but we do need to be reminded of the potential negative effects of habit. Nevertheless, these considerations do not falsify the claim that there is an element of habit in virtue. So far, then, our analysis of the components of intellectual virtue has identified a component of habitual motivation arising from the motive to know and a

component of reliable success in achieving the aims of the virtue in question.

I have said that the primary motivation underlying the intellectual virtues is the motivation for knowledge. Such a motivation clearly includes the desire to have true beliefs and to avoid false ones, and we have looked at how such a motivation leads a person to follow rules or procedures of belief formation that are known to her epistemic community to be truth conducive. The motivation for knowledge also leads its possessor to acquire the motivational components distinctive of the individual intellectual virtues: open-mindedness, fair-mindedness, intellectual flexibility, and so on. And the motivation to be, say, open-minded, will lead to acquiring patterns of behavior characteristic of the open-minded; the motivation to be fair-minded will lead to acquiring patterns of behavior characteristic of the intellectually fair; and so on. It is doubtful that such patterns of behavior are fully describable in terms of following rules or procedures. It is clear, then, that the following of truth-conducive procedures is not all that a knowledge-motivated person does, both because the motivation for truth leads to behavior that is not fully describable as the following of procedures, and because the motivation for knowledge includes more than the motivation for truth. The motivation for knowledge leads us to be aware of the reliability of certain belief-forming processes and the unreliability of others, but it also leads us to be aware that there are reliable belief-forming mechanisms whose reliability is not yet known. And similarly, there are unreliable belief-forming mechanisms whose unreliability is not yet known. This is something we cannot ignore; otherwise, knowledge about knowledge would not progress. This means that intellectual virtues such as flexibility, open-mindedness, and even boldness are highly important. It also suggests that there is more than one sense in which a virtue can be truth conducive. In the sense most commonly discussed by reliabilists, truth conduciveness is a function of the *number* of true beliefs and the *proportion* of true to false beliefs generated by a process. There is another sense of truth conduciveness, however, which is important at the frontiers of knowledge and in areas, like philosophy, that generate very few true beliefs, no matter how they are formed. I suggest that we may legitimately call a trait or procedure truth conducive if it is a necessary condition for advancing knowledge in some area even though it generates very few true

beliefs and even if a high percentage of the beliefs formed as the result of this trait or procedure are false. For example, the discovery of new reliable procedures may arise out of intellectual traits that lead a person to hit on falsehood many times before hitting on the truth. As long as these traits (in combination with other intellectual virtues) are self-correcting, they will eventually advance human knowledge, but many false beliefs may have to be discarded along the way. A person motivated to know would be motivated to act cognitively in a manner that is truth conducive in this sense, I would argue, in addition to acting in a way that is truth conducive in the more common sense.

The virtues of originality, creativity, and inventiveness are truth conducive in the sense just described. Clearly, their truth conduciveness in the sense of producing a high proportion of true beliefs is much lower than that of the ordinary virtues of careful and sober inquiry, but they are truth conducive in the sense that they are necessary for the advancement of human knowledge. If only 5 per cent of a creative thinker's original ideas turn out to be true, her creativity is certainly truth conducive because the stock of knowledge of the human race has increased through her creativity. The way in which these virtues are truth conducive is probably circuitous and unpredictable, and for this reason it is doubtful that they give rise to a set of rules, and, in fact, they may even defy those rules already established. Often creative people simply operate on intuition, which is usually what we call an ability when it works and we don't know how it works. Ernest Dimnet relates the story that Pasteur was constantly visited by intuitions that he was afterward at great pains to check by the ordinary canons of science (1928, p. 187).¹⁶ Presumably, following the canons in the absence of his bold and original ideas would not have gotten him (or us) nearly as far. Dimnet tells another anecdote about the creative process in novelists. Apparently, when Sir Walter Scott hit upon the idea for a new novel, he would read volume after volume that had no reference to his subject, merely because reading intensified the working of his mind. Dimnet comments that this process did for Scott's power of invention what the crowds in the city did for Dickens's (p. 7). Of course, novelists are not aiming for truth in the sense that is the major focus of this book, but the same point could apply to creative work in philosophy, history, mathematics, and the sciences. The

knowledge-motivated person will want to have the virtues of creativity to the extent that she is able, and that gives us another reason why the motive to know includes more than the motive to follow procedures known to be reliable. The division of epistemic labor probably limits the number of people who are strongly motivated in this way, but their existence is important for the knowledge of the whole community.¹⁷

In "The Doctrine of Chances," C. S. Peirce expressed the opinion that even the scientific method is truth conducive only in a sense similar to the one I have just described. Peirce says that the scientist must be unselfish because he is not likely to arrive at the truth for himself in the short run. Instead, his procedures are likely to lead the scientific community to better theories and more comprehensive truths in the long run.¹⁸ If Peirce is right, the sense in which the virtues of originality and creativity are truth conducive is not clearly different from the way in which the virtues of careful scientific inquiry are truth conducive.

Another reason the motivation to know is not fully expressed by following well-known reliable belief-forming processes is that, as already remarked, the motivation to know includes the motivation for understanding. Knowledge has been associated with certainty and understanding for long periods of its history, but generally not

with both at the same time.¹⁹ The virtues that lead to the kind of knowledge that gives the possessor certainty may be different from the virtues that lead to understanding, and the following of belief-forming processes known to be reliable may be insufficient for either one. For one thing, to aim at certainty is not just to aim at truth but to aim to have an awareness of truth that has a certain quality. To get an awareness with that quality it may not be enough to use processes known or truly believed by one's epistemic community to be reliable. One may need to be aware of how and why one's belief-forming process is justified, or at least how and why it is reliable and the degree of its reliability. The virtues that enable one to see how one's belief can stand up to attack contribute to certainty. Virtues that lead to clarity in one's grasp of a matter may also contribute to certainty. Aiming at understanding is even farther removed from using procedures known to be reliable, because understanding is not a property whose object is a single proposition. Those virtues that enable the agent to see connections among her beliefs – introspective attentiveness and insight in its various forms – are understanding conducive. All of these virtues deserve careful attention, and although I will not stop to investigate them individually, I hope that others will do so.

Notes

- 1 Alisdair MacIntyre, *After Virtue* (Notre Dame, IN: University of Notre Dame Press, 1984).
- 2 Gregory Trianosky, "Virtue, Action, and the Good Life: Towards a Theory of the Virtues," *Pacific Philosophical Quarterly* 68 (1987), pp. 124–47.
- 3 Robert C. Roberts, "Aristotle on Virtues and Emotions," *Philosophical Studies* 56 (1989), pp. 293–306.
- 4 In *Reason and the Heart: A Prolegomenon to a Critique of Passional Reason* (Ithaca, NY: Cornell University Press, 1996), ch. 2, William Wainwright discusses the love of being in general as an epistemic virtue recognized by Jonathan Edwards.
- 5 Hobbes implies here that people do not differ as much in their sensory faculties as in their virtues and vices. He also says that part of what leads us to call a quality a virtue is that it is uncommon. The Hobbesian approach would hesitate, then, in attributing anything virtuous to cases of simple perceptual beliefs that are produced by normally functioning faculties.
- 6 Thomas Hobbes, *Leviathan* (New York: Macmillan, 1958), pt. 1, ch. 8, pp. 68–9.
- 7 John Dewey, *How We Think* (Boston: D. C. Heath and Co., 1933), ch. 2, sec. 1; page numbers given in parenthesis in the text.
- 8 Laurence Bonjour, "Externalist Theories of Empirical Knowledge," in *Studies in Epistemology*. Midwest Studies in Philosophy, vol. 5 (Notre Dame, IN: Notre Dame University Press, 1980); Hilary Kornblith, "Justified Belief and Epistemically Responsible Action," *Philosophical Review* 92 (1983), pp. 33–48.
- 9 James A. Montmarquet, "Epistemic Virtue," *Mind* 96 (1986), pp. 482–97; "Epistemic Virtue," in Jonathan Dancy and Ernest Sosa (eds), *A Companion to Epistemology* (Oxford: Basil Blackwell, 1992); *Epistemic Virtue and Doxastic Responsibility* (Lanham, MD: Rowman and Littlefield, 1993), ch. 2.
- 10 *Epistemic Virtue*, pp. 27–8.
- 11 *Ibid.*, pp. 26–7.

- 12 Charles Young has suggested to me that a problematic case is the desire that the interesting be true, a quality whose entrenchment might have value, he suggests, independently of its capacity to reliably lead to the truth. There is a passage in the *Meno* (81de) in which Socrates seems to be saying that even if we have no rational grounds for preferring the religious story of 81ad to the eristic story of 80d, we are better off believing the former: it makes us energetic seekers, whereas the eristic story makes us lazy. I have noticed in myself and others the tendency to go for the more metaphysically exciting position on such issues as the nature of time or the existence of abstract objects, quite apart from a consideration of the weight of the argumentative evidence. Such a tendency is clearly dangerous, but it is not obviously a bad thing. There might even be value in its entrenchment.
- 13 I thank Hilary Kornblith for drawing my attention to this point.
- 14 I thank Charles Young for this point.
- 15 Amelie Rorty, "From Exasperating Virtues to Civic Virtues," *American Philosophical Quarterly* 38 (1996), pp. 303–14.
- 16 Ernest Dimnet, *The Art of Thinking* (New York: Simon and Schuster, 1928), p. 187.
- 17 A careful study of the psychology of creativity would probably show that motivation operates in a different way in the virtues of creativity and originality than it does in the other intellectual virtues. The motivation to be creative does not lead to being creative in the way the motivation to be careful leads to being careful. I imagine that creative people begin by being creative involuntarily and find it pleasant, exciting, even thrilling. These feelings give them the impetus to permit their creativity a certain latitude, which may lead them to ignore the established canons, at least temporarily. This means that the motivational component in creativity does not so much lead its possessors to *acquire* the trait as allow them to give it free rein, and this may lead to ignoring the dictates of certain other virtues.
- 18 Christopher Hookway has an interesting discussion of this position of Peirce and related views. See C. S. Peirce, *The Essential Peirce*, vol. 1, ed. Nathan Houser and Christian Kloesel (Bloomington: Indiana University Press, 1992) and Christopher Hookway, "Mimicking Foundationalism: On Sentiment and Self-Control," *European Journal of Philosophy* 1:2 (1993), pp. 156–74.
- 19 See Mary Tiles and Jim Tiles, *An Introduction to Historical Epistemology* (Cambridge, MA: Basil Blackwell, 1993), and Stephen Everson, *Epistemology, Companions to Ancient Thought*, vol. 1 (Cambridge: Cambridge University Press, 1990), for historical discussions of the difference between the values of certainty and understanding in different periods of epistemological history.

Virtues and Vices of Virtue Epistemology

John Greco

In this paper I want to examine the virtues and vices of virtue epistemology. My conclusion will be that the position is correct, when qualified appropriately. The central claim of virtue epistemology is that, Gettier problems aside, knowledge is true belief which results from a cognitive virtue. In section one I will clarify this claim with some brief remarks about the nature of virtues in general, and cognitive virtues in particular. In section two I will consider two objections to the theory of knowledge which results. In section three of the paper I will argue that virtue epistemology can be qualified so as to avoid the objections raised in section two.

Specifically, I will argue that not all reliable cognitive virtues give rise to knowledge. Rather, a cognitive virtue gives rise to knowledge only if (i) it is reliable, and (ii) the reliability of the virtue is the result of epistemically responsible doxastic practices. In cases of knowledge, reliability is grounded in responsible belief formation and maintenance. The resulting position has ramifications for the analysis of knowledge, the internalism-externalism debate concerning epistemic justification, and the problem of skepticism.

I What is a Cognitive Virtue?

A virtue, in one important sense, is an ability. An ability, in turn, is a stable disposition to achieve certain results under certain conditions. Further, when we say that a subject *S* has an ability to

achieve certain results, we imply that it is no accident that *S* achieves those results. *S*'s disposition to achieve the relevant results is grounded in certain properties of *S*, such that under the appropriate conditions any subject with those properties would tend to achieve those results.

For example, Don Mattingly has the ability to hit baseballs. This means that Mattingly has a stable disposition to hit baseballs under appropriate conditions, although Mattingly will not hit the baseball every time under those conditions. Further, it is no accident that Mattingly tends to hit baseballs. Mattingly's tendency to hit baseballs is grounded in certain properties of Mattingly, such that anyone with those properties would also tend to hit baseballs with similar success in similar conditions.

A more exact definition of cognitive virtue is as follows:

- (V) A mechanism *M* for generating and/or maintaining beliefs is a cognitive virtue if and only if *M* is an ability to believe true propositions and avoid believing false propositions within a field of propositions *F*, when one is in a set of circumstances *C*.

According to the above formulation, what makes a cognitive mechanism a cognitive virtue is that it is reliable in generating true beliefs rather than false beliefs in the relevant field and in the relevant circumstances. It is correct to say, therefore, that virtue epistemology is a kind of reliabilism. Whereas generic reliabilism maintains that justified belief is belief which results from a reliable cognitive process, virtue epistemology puts a

restriction on the kind of process which is allowed. Specifically, the cognitive processes which are important for justification and knowledge are those which have their bases in a cognitive virtue.

Let us use the term 'positive epistemic status' to designate that property (whatever it may be) which turns true belief into knowledge, Gettier problems aside. Then an important corollary of virtue epistemology is as follows.

(VE) S's belief that p has positive epistemic status for S if and only if S's believing that p is the result of some cognitive virtue of S.

The claim embodied in (VE) has a high degree of initial plausibility. By making the idea of faculty reliability central, virtue epistemology explains nicely why beliefs caused by perception and memory often have positive epistemic status, while beliefs caused by wishful thinking and superstition do not. Second, the theory gives us a basis for answering certain kinds of skepticism. Specifically, we may agree that if we were brains in a vat, or victims of a Cartesian demon, then we would not have knowledge even in those rare cases where our beliefs turned out true. But virtue epistemology explains that what is important for knowledge is that our cognitive faculties are *in fact* reliable in the conditions we are in. And so we do have knowledge so long as we are in fact *not* victims of a Cartesian demon, or brains in a vat.

But although virtue epistemology has initial plausibility, it faces at least two substantial objections. I turn to those objections now.

II Objections to Virtue Epistemology

1 *The evil demon problem for virtue epistemology*

The first objection faced by virtue epistemology is that (VE) seems too strong. This objection arises if we think that positive epistemic status is closely related to epistemic justification. More specifically, it seems possible that an epistemic agent could be justified in believing that p , even when her intellectual faculties are largely unreliable. Suppose, for example, that Kathy is the victim of a Cartesian deceiver. Despite her best efforts almost none of Kathy's beliefs about the world around her are true. It is clear that in this case Kathy's faculties of perception are almost wholly

unreliable. But we would not want to say that none of Kathy's perceptual beliefs are justified. If Kathy believes that there is a tree in her yard, and if she bases this belief on the kind of experience usually caused by trees, then it seems that she is as justified as we would be regarding a similar belief. The problem for virtue epistemology is to account for this intuition. There is something about Kathy's belief which is epistemically valuable, i.e. valuable in a way which is relevant for having knowledge. Yet it is clear that Kathy's belief is not the result of a cognitive virtue in the sense defined by (V).

Sosa's strategy for addressing the evil demon problem is to make justification relative to an environment. Thus Sosa recognizes that there is something valuable about Kathy's belief, even though that belief has its origin in wholly unreliable cognitive faculties. What is valuable about Kathy's belief, Sosa argues, is that it is produced by cognitive faculties which would be reliable in *our* environment.

On the present proposal, aptness is relative to an environment. Relative to our actual environment A, our automatic experience-belief mechanisms count as virtues that yield much truth and justification and aptness. Of course, relative to the demonic environment D, such mechanisms are not virtuous and yield neither truth nor aptness. It follows that relative to D the demon's victims are not apt, and yet *relative to A their beliefs are apt.*¹

The above proposal by Sosa is an interesting one, but some questions arise. First, couldn't we construct the example so that Kathy's cognitive mechanisms are *not* reliable relative to our environment? Thus suppose that Kathy is a brain in a vat, hooked up to a super computer which causes her to have experiences exactly similar to the experiences that I am having now. If in these circumstances Kathy forms the belief that there is a glass of water on the table in front of her, her belief should be as justified as is my belief that there is a glass of water on the table in front of me. But if Kathy were in my environment her cognitive faculties would not be reliable at all, and in fact would be incapable of connecting her with reality at all. For if Kathy were in my environment, rather than hooked up to a super computer, she would lack the faculties for producing experiences. She would be a helpless brain on a desk.

Or suppose that Kathy's powers of reasoning are *helped* by her vat environment. We may imagine that Kathy's natural reasoning mechanisms are defective, but that the fluids in the vat serve to correct the defect. Thus inside the vat environment Kathy is a flawless reasoner. But if Kathy were in a normal environment, i.e., inside a normal head with normal sensory apparatus, her reasoning mechanisms would be defective and thus unreliable. Now suppose that Kathy believes that the house in front of her was built before 1900, and that she believes this partly on the basis of her present experience and partly on the basis of her reasoning from this experience. It seems to me that Kathy could be perfectly justified in this belief, even though she is a brain in a vat and the cognitive mechanisms which produce her belief are not reliable. However, the mechanisms which produce Kathy's belief are not reliable relative to our environment either, since in our environment Kathy would lack the vat fluids which correct her cognitive defects.

Sosa might attempt to solve the problem as follows. We could define environments very specifically, so as to include being in a normal head, etc., and then define sets of circumstances in terms of experiential and doxastic inputs. Cognitive mechanisms would then be dispositions to form certain beliefs in a field, given certain experiential and doxastic inputs. We could then say that what is valuable about Kathy is not that she would *have* virtues if she were in our environment, but that the cognitive mechanisms she does have in her environment would *be* virtues in our environment.²

I take it that there are at least two problems with the latest proposal: one for Sosa's positions in particular; and one for virtue epistemology in general. First, Sosa's position requires an epistemic perspective on one's own cognitive virtues in order to have reflective justification, and in order to solve the generality problem.³ But it is implausible that the typical believer has such a perspective when virtues are defined in terms of experiential inputs. I take it that the experience-belief pairs that would describe a reliable mechanism must be very detailed regarding the quality of the experiences involved. But in the typical case there is no such detailed perspective on even our present experiences, much less the range of our possible experiences.

Second, Sosa's account of what is valuable about Kathy's belief assumes that Kathy's cognitive mechanisms would be reliable in our environment.

But this assumes that *our* cognitive mechanisms are reliable in our environment. We think Kathy is reliable because she is like us in relevant respects, and we think we are reliable. But suppose we are victims of an evil deceiver, or that we are brains in a community vat. Then Kathy's mechanisms are no more reliable in our environment than they are in hers. And thus, according to Sosa's account, Kathy's beliefs are not justified relative to our environment. But this seems wrong – there seems to be something valuable about Kathy's beliefs whether or not she or we are victims of an evil deceiver. There is something epistemically important about the way her and our beliefs are formed, whether or not they are formed via cognitive faculties which are objectively reliable relative to our environment.

We can pursue this point by considering the second of our two objections to virtue epistemology.

2 *The problem of epistemic irresponsibility*

The second objection to be considered is that (VE) is too weak. Specifically, we can imagine cases where S's cognitive faculties are highly reliable with respect to his belief that *p*, but where S is epistemically irresponsible in believing that *p*. Such a case may arise when S has substantial but misleading evidence against his belief that *p*.

Consider the case of Mary, who is in most respects a normal human being. The relevant difference is that Mary's cognitive faculties produce the belief in her that there is a tiger nearby whenever there is a tiger nearby, and even in cases where Mary does not see, hear or otherwise perceive a nearby tiger. Mary's brain is designed so as to be sensitive to an electromagnetic field emitted only by tigers, thus causing her to form the relevant belief in the appropriate situation, and without any corresponding experience, sensory or otherwise. We can imagine that this cognitive feature was designed by natural processes of evolution, or that it was literally designed by a beneficent creator, one who realizes that tigers are dangerous to beings like Mary and who therefore wishes to equip her with a reliable warning device. Now suppose that a tiger is walking nearby, and that Mary forms the appropriate belief. Add that Mary has no evidence that there is a tiger in the area, nor any evidence that she has such a faculty. Rather, she has considerable evidence

against her belief that there are tigers in the area. Clearly, Mary's belief that there is a tiger nearby does not have positive epistemic status in this situation, even though the belief is caused by properly functioning faculties in an appropriate environment. Mary does not *know* that there is a tiger nearby. Again, the explanation for this is that Mary's belief is epistemically irresponsible. Given the way things look from Mary's point of view, she *ought not* to believe that there is a tiger nearby.

Sosa's strategy for addressing this kind of example recognizes the importance of S's point of view by invoking S's epistemic perspective. Sosa makes a distinction between animal knowledge and reflective knowledge. For animal knowledge, it is sufficient that S's true belief be caused by a reliable faculty. For reflective knowledge, we must add that S has a true grasp of the fact that her belief is grounded in a reliable cognitive faculty. This grasp must in turn result from a faculty of faculties, which gives rise to the required epistemic perspective.

For one is able to boost one's justification in favor of *P* if one can see one's belief of *P* as in a field *F* and in circumstances *C*, such that one has a faculty (a competence or aptitude) to believe correctly in field *F* when in conditions *C*. . . . One thereby attributes to oneself some intrinsic state such that when there arises a question in field *F* and one is in conditions *C*, that intrinsic state adjusts one's belief to the facts in that field so that one always or very generally believes correctly.⁴

According to Sosa, to 'see' one's belief that *p* as in a field and circumstances and to 'attribute' to oneself reliability in that field and those circumstances, is to have true beliefs to that effect, where those true beliefs are themselves products of a cognitive virtue. And now we may see how this position can be applied to the case above. According to Sosa, Mary has animal knowledge but not reflective knowledge. Further, he can say that Mary's belief is reflectively unjustified, since her belief actually conflicts with her epistemic perspective on her faculties.

But the problem with this proposal is that we seldom have such beliefs about our beliefs and about our cognitive faculties. In the typical case, we have no beliefs at all about the sources of our beliefs, or about our reliability in particular fields and circumstances.

Or at least this is so for occurrent beliefs. Is it plausible that we typically have such beliefs dispositionally? Where we do have such a dispositional perspective, the field and circumstances that perspective specifies are probably the wrong ones. Specifically, to the extent that I attribute to myself certain cognitive faculties, those faculties are specified much too broadly to be of any use. For example, consider my belief that there is a glass of water on the table. In the typical case I have no occurrent beliefs about the source of this belief in a given cognitive virtue. Let me now consider any dispositional beliefs I might have. After considering the issue for a moment, it occurs to me that my belief about the glass is the result of sight. But if you ask me to get very specific about a field of propositions *F*, or a set of circumstances *C*, such that I am highly reliable in that field when in those circumstances, I am at a loss. I simply do not have very specific beliefs in this area, nor is it plausible that such beliefs are available dispositionally if only I think about it a little more.

We may conclude that in the typical case a believer will not have a true grasp of the inventory of cognitive faculties she possesses, nor will she have a perspective on which faculty is responsible for producing the particular belief in question. On the other hand, there does seem to be something importantly right about Sosa's proposal. I want to argue that Sosa is right to invoke S's point of view as an important element for having knowledge, but that he invokes S's point of view in the wrong sense. Below I will develop a different sense in which Mary's belief is correct or appropriate from her point of view, and I will argue that this is the sense which is relevant for having knowledge.

III An Internalist Version of Virtue Epistemology

We have said that (VE) fails to take into account an important kind of epistemic value. Namely, (VE) fails to recognize the importance of S's belief being correct or appropriate from S's point of view. One way in which this lack presents itself is in the evil demon problem. Virtue epistemology fails to recognize an appropriate sense in which S's beliefs might be epistemically valuable, even if those beliefs result from wholly unreliable cognitive faculties. Another way in which this problem presents itself is in examples which show that (VE) is

too weak. There are cases where S's true belief is the result of a reliable cognitive faculty, but where S lacks knowledge because S's belief is somehow inappropriate from S's point of view.

Sosa tries to address the latter problem by invoking the idea of an epistemic perspective. The problem with this proposal is that the relevant perspective is lacking in the typical case. It is implausible that believers typically have a true grasp of their cognitive faculties, or a true grasp of which faculty has produced a particular belief. Below I want to develop a different understanding of what it means for a belief to be correct or appropriate from S's point of view. I will then argue that virtue epistemology can be amended so as to incorporate this understanding, and I will defend the theory of knowledge which results.

1 Norm internalism

Norm internalism is the position that justified belief is the result of following correct epistemic norms, or correct rules of belief formation and maintenance. More exactly,

- (NI) S is epistemically justified in believing that p if and only if S's believing that p is in conformance with the epistemic norms which S countenances, and the history of S's belief has also been in conformance with those norms.⁵

It will be necessary to say more about two of the central notions involved in (NI): the notion of a belief being in conformance with an epistemic norm, and the notion of an epistemic norm being countenanced. I begin with the latter.

We may get an idea of what it is to countenance an epistemic norm if we consider the following example. Suppose that Jane, who is not very good at math, bases her belief in a complicated theorem on a set of axioms which do in fact support the theorem. But suppose that she does so not because she sees the supporting relation, but because she has reasoned invalidly from the axioms to the theorem. Obviously Jane is not justified in her belief that the theorem is correct.

What is required for Jane to be justified in believing the theorem? What is it that justification requires but Jane lacks? A plausible suggestion is that Jane must be sensitive to the inference relation

between her theorem and the axioms on which she bases the theorem. Just what this sensitivity amounts to, however, is not easy to state. For although we are often 'aware' that some set of evidence supports a conclusion, it is not easy to state what this awareness consists in.

One suggestion is that Jane must believe that her conclusion follows from her evidence. But this is obviously too weak. For Jane could believe that her conclusion follows from her evidence even if she has reasoned fallaciously and has no real insight into how her conclusion follows from her evidence. Alternatively, one might suggest that Jane must believe that the relevant general rule of inference is correct, and that her inference is an instance of the general rule. But this suggestion is too strong. Typically only logicians have beliefs about the deductive rules which govern our reasoning, and it is agreed on all sides that no one has successfully characterized the rules which govern our non-deductive reasoning. But if we typically do not have beliefs about the rules which govern correct reasoning, how are we to understand our sensitivity to such rules?

I suggest that although we do not typically have beliefs about such rules, we do *countenance* such rules in our reasoning. In other words, we follow such rules when we reason conscientiously, although the way in which we follow them does not involve having beliefs about them, either occurrent or dispositional. Thus the way in which we countenance rules of reasoning is analogous to the way we countenance other action-governing norms. The norms which govern good hitting in baseball, for example, are countenanced by good hitters when they are batting conscientiously. But this does not mean that all good hitters are capable of articulating those norms, or otherwise forming true beliefs about them. Not all good hitters make good hitting coaches. In fact, it is possible for a good hitter to form false beliefs about the norms which he countenances when he is actually playing.

So although we do not typically have beliefs about the norms which govern our beliefs, we do countenance certain norms and not others. The norms that we countenance are the norms that we follow when we reason conscientiously. And thus it makes perfect sense to say that someone is reasoning in a way that he does not countenance. This is in fact what happens when we form our beliefs hastily, or fall into wishful thinking, or are swayed by our prejudices.

I now turn to the notion of a belief's being in conformance with a norm. The notion can be made more clear by considering a distinction common in moral philosophy. It is common for moral philosophers to make a distinction between acting in accordance with one's duty and acting for the sake of one's duty. In the former case one's actions happen to coincide with one's duties. In the latter case one's actions are performed *because* one has certain duties. And now a similar distinction can be made with respect to our believings. While some of our beliefs are merely in accordance with the norms of belief formation which we countenance, others of our beliefs are in conformance with those norms in the following sense; they arise, at least partly, *because* we countenance certain norms and not others. The latter beliefs are accepted (at least partly) *because* we follow certain norms when we are reasoning conscientiously.⁶

Notice that the position articulated in (NI) is not subject to the objections raised against Sosa's idea of an epistemic perspective. Thus the position does not require that S have beliefs *about* which norms she countenances, or about which norms are involved in the formation of a particular belief. All that is necessary is that S does in fact countenance the relevant norms, and that S's belief is in fact in conformance with those norms. Second, the present position explains what is valuable about the beliefs of the victim of the evil deceiver. In the case of the evil deceiver, Kathy's beliefs are justified because they are in conformance with the rules of belief formation and maintenance which Kathy countenances. Finally, the account explains why Mary's beliefs do not amount to knowledge. Even though Mary's belief results from a reliable tiger-detecting faculty, Mary's belief is not in accordance with the norms which Mary countenances. Presumably, Mary countenances norms which disallow believing that tigers are present in the absence of any evidence to that effect, or in cases where one has considerable evidence against that belief and no evidence in favor of it.⁷

2 Norm internalism applied to virtue epistemology

We may see how the present position can be used to amend virtue epistemology if we make a distinction between a virtue and the basis for that virtue. We have been understanding virtues as abilities, and we have been understanding abilities

as stable dispositions to achieve certain results under certain conditions. But then the same virtue might have different bases in different subjects. Thus the ability to absorb oxygen into the blood has a different basis in fish than it does in human beings. Similarly, the ability to roll down an inclined plane has a different basis in a pencil than it does in a baseball. Now, according to Sosa, the basis for a cognitive virtue is the inner nature of the cognitive subject. Thus Sosa refers to the subject's inner nature explicitly in his latest account: 'One has an intellectual virtue or faculty relative to an environment E if and only if one has an inner nature I in virtue of which one would mostly attain the truth and avoid error in a certain field of propositions F, when in certain conditions C.'⁸

A different proposal would be that the basis for cognitive virtues, at least where knowledge is concerned, must be S's conformance to the epistemic norms which S countenances. On this proposal cognitive virtues relevant to knowledge are grounded in conscientious belief formation and maintenance, rather than in an unchanging inner nature.

Perhaps the following analogy will clarify the present proposal. Pitching machines and Nolan Ryan both have the ability to throw baseballs at high speeds. But the basis of this virtue in the pitching machine is different from the basis of the virtue in Ryan. Moreover, the basis for the virtue in the machine is the machine's inner nature; given the way that the machine is constructed and given the appropriate conditions, the machine throws baseballs at high speeds. The basis of the same virtue in Ryan is of a different sort; Ryan's ability to throw baseballs is based in Ryan's conformance to the norms governing good throwing. A person might have the same inner nature as Ryan and not have the ability to throw baseballs because that person fails to conform to the proper norms. Consider that Ryan himself would not throw baseballs at high speeds if he did not conform to the norms of good throwing.

The analogy should be obvious. I am suggesting that knowers are more like Ryan than like pitching machines. Specifically, I am suggesting that the virtues associated with knowledge have their bases in conformance to relevant norms rather than in a fixed inner nature.⁹

Applying norm internalism to virtue epistemology results in the following account of positive epistemic status:

- (VEI) S's belief that p has positive epistemic status for S if and only if
- (i) S believes that p ;
 - (ii) S's believing that p is the result of a reliable cognitive virtue V of S; and
 - (iii) S's virtue V has its basis in S's conforming to epistemic norms which S countenances.

3 (VEI) defended

According to (VEI), knowledge is true belief which results from a cognitive virtue, where this virtue has its basis in S's conforming to epistemic norms which S countenances. Thus on the present account, knowledge is virtuous in both a subjective and an objective sense. Knowledge is virtuous in a subjective sense in that knowledge is belief which is correct or appropriate from S's point of view. And this means that in cases of knowledge S's belief is in conformance with the rules of belief formation and maintenance which S countenances. Knowledge is virtuous in an objective sense in that belief which is knowledge is the result of a reliable cognitive faculty. Further, the two ways in which knowledge is virtuous are related. In cases of knowledge, a belief is objectively virtuous *because* it is subjectively virtuous. In other words, in cases of knowledge the basis of S's objectively reliable cognitive virtue is in S's conformance to the epistemic norms which S herself countenances; reliability results from responsibility.

We may now see that (VEI) avoids the two objections raised against virtue epistemology as defined by (VE). Because (VEI) recognizes an internalist element in knowledge, (VEI) explains

what is valuable about the beliefs of the victim of an evil deceiver. Namely, someone whose cognitive faculties are made wholly unreliable by an evil deceiver might nevertheless reason in conformance with the norms that she countenances. Thus the victim of an evil deceiver might have beliefs which are subjectively responsible, even if they are not objectively reliable. Second, (VEI) avoids the counterexamples which show that (VE) is too weak. Specifically, (VEI) *requires* epistemic responsibility for positive epistemic status. And this requires that S's belief be correct or appropriate from S's point of view, in the sense defined by norm internalism.

Thus (VEI) avoids the two objections raised against (VE). Does (VEI) still have the attractive features which we attributed to virtue epistemology at the beginning of the paper? (VEI) continues to explain nicely why beliefs caused by perception and memory often have positive epistemic status, while beliefs caused by wishful thinking and superstition do not. But it has the added advantage of explaining why not all reliable cognitive faculties give rise to positive epistemic status. Thus it explains why Mary's tiger-detecting faculties do not. Second, the theory continues to give us a basis for answering certain kinds of skepticism. Thus it continues to explain why we would lack knowledge *if* we were brains in a vat, or victims of a Cartesian demon, and why we do not lack knowledge so long as this is not the case. (VEI) in fact nuances our answer to skepticism by explaining what is epistemically valuable about the beliefs of victims trapped in the skeptical scenarios.

We may conclude that (VEI) retains all the advantages of (VE), while avoiding problems which (VE) cannot.

Notes

- 1 Ernest Sosa, "Intellectual Virtue in Perspective," in *Knowledge in Perspective: Collected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991), p. 289.
- 2 Here I am indebted to Sosa, who suggested this response in conversation.
- 3 The notion of an epistemic perspective and its role in Sosa's account of reflective justification is discussed below.
- 4 Sosa, *Knowledge in Perspective*, p. 282. Sosa develops this strategy in "Intellectual Virtue in Perspective" and in "Reliabilism and Intellectual Virtue," both in *Knowledge in Perspective*.
- 5 This position is defended in detail in my "Internalism and Epistemically Responsible Belief," *Synthese* 85 (1990), pp. 245–77.
- 6 For a similar distinction, see John Pollock, *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield 1986), p. 168.
- 7 Of course questions remain. For example, it is plausible that a person's norms will change and even conflict over time. How does this effect epistemic responsibility? There is also the problem of "norm schizophrenia," or the problem of conflicting norms at the same time. Finally, it might be thought that the above account leads to an unacceptable kind of epis-

temic relativism. I address all of these questions in my "Internalism and Epistemically Responsible Belief." There I conclude that a) responsibility concerns conformance to one's present norms; b) when present norms conflict responsibility requires that none of S's norms disallow S's belief; and c) the only kind of relativism involved is harmless, and

should be expected given the analogy to moral responsibility.

- 8 Sosa, "Intellectual Virtue in Perspective," p. 284. See also Sosa's definitions on pp. 286–9.
- 9 Perhaps I should say "rather than *merely* in a fixed inner nature," since it is possible that S's conformance to relevant norms is itself based in a deeper inner nature.



PART IX

Epistemic Contextualism

Introduction

The epistemic contextualist's answer to the skeptic is a unique blend: it is a concession and an attribution of error. Consider the basic form of the skeptical argument as formulated by Keith DeRose, with "*H*" standing in for some sentence describing a skeptical hypothesis and "*O*" for a sentence describing an ordinary state of affairs:

1. I don't know that not-*H*
2. If I don't know that not-*H*, then I don't know that *O*. So,
- C. I don't know that *O*.

In reply to the skeptic, the contextualist refuses to deny either of the premises, and indeed does not deny the conclusion. This is the concession. Yet the skeptic is charged with the error of not seeing the implications of the context-sensitivity of knowledge claims, of supposing that knowledge claims are false in all contexts, even in ordinary contexts quite apart from any skeptical challenge. The truth of the matter, says the contextualist, is that the skeptic's dialectical challenge raises the standards for truly attributing knowledge. Still, in many ordinary contexts, lower standards are in place, making true knowledge attributions not only possible but commonplace.

This is the contextualist reply to the skeptic, in broad outline. As is pointed out effectively by DeRose, putting flesh on the bones of this view requires describing the mechanisms by which the standards for truly ascribing knowledge are raised and lowered. The task of providing such descriptions, of course, introduces a danger of generating incorrect predictions regarding the truth-values of knowledge claims in various hypothetical situations. In his contribution, Stewart Cohen argues

that this is the sort of difficulty that undermines David Lewis's recent contextualist account.

Keith DeRose takes as his starting point Robert Nozick's insight that knowledge is importantly connected to sensitivity. Why am I unable to know my lottery ticket is a loser regardless of how large the lottery pool is – unable, that is, *until* I hear the official announcements of the winner? The probability that my ticket is a loser given that it is one of a million (or more) is much *greater* than the probability that the announcer is telling the truth. Nozick answers: before but not after you hear the announcement your belief is insensitive; had your ticket not been a loser, you would have still believed it was.

Nonetheless, DeRose shies away from the pure sensitivity account because it fails to rule out, and even licenses, what he calls "abominable conjunctions." These are conjunctions of the form "I know that *P* but I don't know that not-*Q*," where *P* entails not-*Q*. "Straight" solutions to the skeptical problem, which give up (2) in the argument above, are forced to admit the truth of the abominable "I know that *O* but I don't know that not-*H*." These fly in the face of our intuitions, and should be regarded as false in every context.

DeRose is therefore constrained not to locate the source of the context-dependence in the conditional premise, (2). He locates the source, instead, in the assertion of (1) itself. The effect of denying knowledge is to require thereafter a stronger epistemic position for knowledge than was required before. It is this notion of *strength of epistemic position* that receives an account in terms of sensitivity. We judge relative strength of epistemic positions when we judge the truth-values of

conditionals of the form “If I know that P , then I know that Q ,” e.g., “If I know that I own a cat, then I know that I own an animal,” as well as those of their converses. What guides us in judging such conditionals true or false, moreover, are facts about the distance from actuality (if any) that would be required for the belief/fact link to be broken. Thus, a cat-owner S ’s epistemic position with respect to “I know I own an animal” will be stronger than his epistemic position with respect to “I know I own a cat,” since one would have to look farther from actuality to find a possible world in which S believes falsely that he has a cat than one would to find a world in which S believes falsely that he has an animal. Sensitivity then enters as a limit to which the standards for strength of epistemic position can be raised. This is precisely the contextual feature that the skeptic exploits. By claiming that one lacks knowledge, say, that one is not a brain in a vat, the skeptic raises the contextual standards so high as to require sensitivity to the truth of the proposition that one is not a brain in a vat. But once the sphere of relevant worlds is so far extended, one’s belief that one has hands will not match the facts across all the worlds. One’s epistemic position with respect to one’s belief that one has hands, after all, is not as strong as it is with respect to one’s belief that one is not a brain in a vat. (This is a noteworthy claim in light of the fact that Dretske, Nozick, and others insist that while we may know that we have hands, we certainly do not know that we are not brains in vats.)

David Lewis’s contextualism takes off from the intuitive infallibilist claim that one knows that p just in case one’s evidence eliminates all possibilities in which not- p . This does not lead to skepticism (in all contexts), however, because the scope of “all” is taken to be contextually restricted. Thus, Lewis adds to this biconditional, *sotto voce*, “– Psst! – except for those possibilities that we are properly ignoring.” A significant portion of Lewis’s paper is then devoted to listing rules that determine at least partially what can and cannot be properly ignored. Corresponding to the traditional triad of requirements for knowledge, truth, belief, and justification, we encounter a pair of rules: the actuality and the belief rules. If a possibility is actual (believed by the subject to be actual), it cannot be properly ignored. Prominent among the other rules listed are the rules of resemblance and attention. The rule of resemblance states that if a possibility saliently resembles another possibil-

ity, then if one cannot be properly ignored neither can the other. The rule of attention states that if a possibility is attended to, it cannot be properly ignored. This framework is then applied to three kinds of problem: skepticism, the lottery problem, and the Gettier problem. The application to skepticism appeals to the rule of attention. To truly claim that S knows that p after the mention of the appropriate skeptical possibility, S ’s evidence would have to eliminate the skeptical possibility attended to, which it will not do. Applied to lottery and Gettier cases, the rules of actuality and resemblance come to the forefront. Since the possibility in which S has the winning ticket is saliently similar to actuality (whatever actuality may be), the possibility of having a winner is not properly ignored. Thus, the knowledge attribution will be false. Similarly, since the possibility in which neither Nogot nor Havit owns a Ford saliently resembles actuality, it is not properly ignored. Here, too, the knowledge attribution will be false.

Stewart Cohen argues that Lewis’s rules wrongly assimilate the Gettier problem to the lottery problem and skepticism. The salience of resemblance requirement is appropriate in regard to the latter cases, but not the former. To establish this, Cohen first distinguishes between speaker-sensitive and subject-sensitive rules for properly ignoring possibilities, noting that the rule of salient resemblance is both speaker- and subject-sensitive. He then examines the consequences. Suppose S sees a sheep-shaped rock on a hill, and taking the rock to be a sheep comes to believe that there is a sheep on the hill. As it turns out, there is in fact a sheep behind the rock, out of S ’s view. Now an attributor who knows all these facts will find the possibility that S sees a sheep-shaped rock on a sheepless hill salient, but one who views the scene from a different angle may not realize that S is in fact seeing a rock. The second attributor will not find salient the possibility just mentioned, nor any other possibility like it. In fact, then, the second attributor can truly attribute knowledge to S . Since facts about salience vary from attributor to attributor, facts about whether it’s true to say of someone in a Gettier case that she has knowledge, too, will vary.

Cohen argues that this result is implausible in the face of our stable disposition to deny knowledge to subjects in Gettier cases. A contextualist theory should appeal to speaker-sensitive rules only when our intuitions vacillate. We do vacillate about the truth of skeptical claims and about the

truth of knowledge attributions in certain lottery cases, such as that of Lewis's Poor Bill: "Pity poor Bill! He squanders all his spare cash on the pokies, the races, and the lottery. He will be a wage slave

all his days. We know he will never be rich. But if he wins the lottery (if he wins big), then he will be rich." We don't vacillate about Gettier cases.

Further Reading

- Annis, David, "A Contextualist Theory of Epistemic Justification," *American Philosophical Quarterly* 15 (1978), pp. 213–19.
- Austin, J. L., "Other Minds," in his *Philosophical Papers*, 3rd edn (Oxford: Clarendon Press, 1979), pp. 76–116.
- Cohen, Stewart, "Skepticism and Everyday Knowledge Attributions," in Roth and Ross (eds), *Doubting: Contemporary Perspectives on Skepticism* (Dordrecht: Kluwer Academic Publishers, 1990), pp. 161–9.
- , "How to be a Fallibilist," *Philosophical Perspectives* 2 (1988), pp. 91–123.
- DeRose, Keith, "Contextualism and Knowledge Attributions," *Philosophical and Phenomenological Research* 52 (1992), pp. 913–29.
- Dretske, Fred, "Epistemic Operators," *Journal of Philosophy* 67 (1970), pp. 1007–23.
- Hambourger, Robert, "Justified Assertion and the Relativity of Knowledge," *Philosophical Studies* 51 (1987), pp. 241–69.
- Kaplan, Mark, "Epistemology on Holiday," *Journal of Philosophy* 88 (1991), pp. 132–54.
- Lewis, David, "Scorekeeping in a Language Game," *Journal of Philosophical Logic* 8 (1979), pp. 339–59, esp. Example 6, "Relative Modality," pp. 354–5.
- Schiffer, Stephen, "Contextualist Solutions to Scepticism," *Proceedings of the Aristotelian Society* 96 (1996), pp. 249–61.
- Sosa, Ernest, "Chisholm's Epistemology and Epistemic Internalism," in Lewis Hahn (ed.), *The Philosophy of Roderick Chisholm* (Chicago: Open Court, 1997).
- , "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999).
- , "Contextualism and Skepticism," *Philosophical Issues* 10 (1999).
- Unger, Peter, *Philosophical Relativity* (Minneapolis: University of Minnesota Press, 1984).
- , "The Cone Model of Knowledge," *Philosophical Topics* 14 (1986), pp. 125–78.
- Wittgenstein, Ludwig, *On Certainty* (New York: Harper and Row, 1972).
- Wright, Crispin, "Dreaming and Scepticism: Imploding the Demon," *Mind* 100 (1991), pp. 87–116.

Solving the Skeptical Problem

Keith DeRose

1 The Puzzle of Skeptical Hypotheses

Many of the most celebrated, intriguing, and powerful skeptical arguments proceed by means of skeptical hypotheses. Brutally pared to their barest essentials, they are roughly of the following form, where "O" is a proposition about the external world one would ordinarily think one knows (e.g., I have hands¹) and "H" is a suitably chosen skeptical hypothesis (e.g., I am a bodiless brain in a vat who has been electrochemically stimulated to have precisely those sensory experiences I've had, henceforth a "BIV"²):

*The Argument from Ignorance (AI)*³

1. I don't know that not-H.
2. If I don't know that not-H, then I don't know that O. So,
- C. I don't know that O⁴

Setting aside the distracting side issues that immediately threaten from all directions, and keeping AI in this stark, uncomplicated form, I will, in what follows, present and defend, at least in broad outline, the correct solution to the puzzle AI confronts us with. And AI does present us with a puzzle, because, for reasons we'll investigate in later sections, each of its premises is initially plausible, when H is well chosen. For however improbable or even bizarre it may seem to suppose that I am a BIV, it also seems that I don't know that I'm not one. How could I know such a thing? And it

also seems that if, for all I know, I am a BIV, then I don't know that I have hands. How could I know that I have hands if, for all I know, I'm bodiless (and therefore handless)? But, at the same time, it initially seems that I do know that I have hands. So two plausible premises yield a conclusion whose negation we also find plausible. So something plausible has to go. But what? And equally importantly, how?

To be sure, the premises are only plausible, not compelling. Thus, we will always have recourse to the Moorean reaction to this argument: Declare that it is more certain that one knows that one has hands than it is that either of the premises of the argument is true (much less that their conjunction is true), and therefore reject one of those premises, rather than accept the conclusion. But also available is the skeptical reaction, which is to accept the conclusion.

But we should hope for a better treatment of the argument than simply choosing which of the three individually plausible propositions — the two premises and the negation of the conclusion — seems least certain and rejecting it on the grounds that the other two are true. In seeking a solution to this puzzle, we should seek an explanation of how we fell into this skeptical trap in the first place, and not settle for making a simple choice among three distasteful ways out of the trap. We must explain how two premises that together yield a conclusion we find so incredible can themselves seem so plausible to us. Only with such an explanation in place can we proceed with confidence and with understanding to free ourselves from the trap.

Many of those working on AI in recent years seem to have understood this.⁵ And I have good

Originally published in *The Philosophical Review* 104, 1 (1995), pp. 1-7, 17-52; copyright Cornell University, reprinted by permission of the publisher and the author.

news to report: Substantial progress towards finally solving this skeptical puzzle has been made along two quite different fronts. The bad news is that, as I shall argue, neither approach has solved the puzzle. But the culminating good news is that, as I will also argue, the new solution I present here, which incorporates important aspects of each of the two approaches, *can* finally solve this perennially thorny philosophical problem. While more details and precision will be called for in the resulting solution than I will provide, there will be enough meat on the bones to make it plausible that the fully articulated solution lies in the direction I point to here.

In sections 2–4 of this paper, I explore the contextualist approach to the problem of skepticism, and show why it has thus far fallen short of solving the puzzle. In sections 3–7, I turn to Robert Nozick's attempt to solve our puzzle. Since the shortcomings of Nozick's treatment of knowledge and skepticism have been, at least to my satisfaction, duly demonstrated by others, it will not be my purpose here to rehearse those shortcomings, but rather to explore and expand upon the substantial insight that remains intact in Nozick's account. In sections 8–14, I present and defend my own contextualist solution, which I argue is the best solution to our puzzle. Since, as I argue in sections 12–14, the skeptic's own solution, according to which we accept AI's conclusion, is among the solutions inferior to the one I present, AI does not successfully support that conclusion.

2 Contextualist Solutions: The Basic Strategy

Suppose a speaker A (for "attributor") says, "S knows that *P*," of a subject S's true belief that *P*. According to contextualist theories of knowledge attributions, how strong an epistemic position S must be in with respect to *P* for A's assertion to be true can vary according to features of A's conversational context.⁶

Contextualist theories of knowledge attributions have almost invariably been developed with an eye toward providing some kind of answer to philosophical skepticism. For skeptical arguments like AI threaten to show, not only that we fail to meet very high requirements for knowledge of interest only to misguided philosophers seeking absolute certainty, but that we don't meet even the truth con-

ditions of ordinary, out-on-the-street knowledge attributions. They thus threaten to establish the startling result that we never, or almost never, truthfully ascribe knowledge to ourselves or to other mere mortals.

But, according to contextualists, the skeptic, in presenting her argument, manipulates the semantic standards for knowledge, thereby creating a context in which she can *truthfully* say that we know nothing or very little.⁷ Once the standards have been so raised, we *correctly* sense that we only could falsely claim to know such things as that we have hands. Why then are we puzzled? Why don't we simply accept the skeptic's conclusion and henceforth refrain from ascribing such knowledge to ourselves or others? Because, the contextualist continues, we also realize this: As soon as we find ourselves in more ordinary conversational contexts, it will not only be true for us to claim to know the very things that the skeptic now denies we know, but it will also be wrong for us to deny that we know these things. But then, isn't the skeptic's present denial equally false? And wouldn't it be equally true for us now, in the skeptic's presence, to claim to know?

What we fail to realize, according to the contextualist solution, is that the skeptic's present denials that we know various things are perfectly compatible with our ordinary claims to know those very propositions. Once we realize this, we can see how both the skeptic's denials of knowledge and our ordinary attributions of knowledge can be correct.

Thus, it is hoped, our ordinary claims to know can be safeguarded from the apparently powerful attack of the skeptic, while, at the same time, the persuasiveness of the skeptical argument is explained. For the fact that the skeptic can invoke very high standards that we don't live up to has no tendency to show that we don't satisfy the more relaxed standards that are in place in more ordinary conversations and debates.

Three important points about contextualist strategies as described above should be made before I move on. First, this type of strategy will leave untouched the timid skeptic who purports by AI merely to be establishing the weak claim that in *some* (perhaps "high" or "philosophical") sense (perhaps induced by the presentation of AI) we don't know the relevant *O*, while not even purporting to establish the bold thesis that our ordinary claims to know that same proposition are false. Whether such a timid skeptical stance is of any

interest is a topic for another paper. The contextualist strategy is important because AI initially seems to threaten the truth of our ordinary claims – it threatens to boldly show that we’ve been wrong all along in thinking and saying that we know this and that. For it doesn’t seem as if it’s just in some “high” or “philosophical” sense that AI’s premises are true: They seem true in the ordinary sense of “know.” In fact, one is initially tempted to say that there’s *no* good sense in which I know that I’m not a BIV or in which I can know I have hands if I don’t know that I’m not a BIV. How (and whether) to avoid the bold skeptical result is puzzle enough.

Second, in presenting the contextualist strategy, I have above assumed a skeptic-friendly version of contextualism – one according to which the philosophical skeptic can (fairly easily), and does, succeed in raising the standards for knowledge in such a way as to make her denials of knowledge true. Some contextualists may think that it’s not so easy to so raise the standards for knowledge, and that a determined opponent of the skeptic can, by not letting the skeptic get away with raising them, keep the standards low. But the important point is to identify the mechanism by which the skeptic at least threatens to raise the standards for knowledge. Whether the skeptic actually succeeds against a determined opponent in so raising the standards is of little importance. To safeguard ordinary claims to know while at the same time explaining the persuasiveness of the skeptical arguments (which is the goal of his strategy), the contextualist can provisionally *assume* a skeptic-friendly version of contextualism, leaving it as an open question whether and under which conditions the skeptic actually succeeds at raising the standards. The contextualist’s ultimate point will then be this: To the extent that the skeptic *does* succeed, she does so only by raising the standards for knowledge, and so the success of her argument has no tendency to show that our ordinary claims to know are in any way defective.

Third, AI can be puzzling even when one is not in the presence of a skeptic who is presenting it. The argument has about the same degree of intuitive appeal when one is just considering it by oneself, without anybody’s *saying* anything. But the contextualist explanation, as described above, involves the standards for knowledge being changed by what’s being said in a conversation.⁸ For the most part, I will frame the contextualist explanation in terms of such conversational rules,

largely because that’s what been done by my contextualist predecessors, with whom I want to make contact. But we must realize that the resulting solution will have to be generalized to explain why the argument can be so appealing even when one is considering it in solitude, with nothing being said. The basic idea of the generalization will take either or both of the following two forms. First, it can be maintained that there is a rule for the changing of the standards for knowledge that governs the truth conditions of our *thoughts* regarding what is and is not known that mirrors the rule for the truth conditions of what is *said* regarding knowledge. In that case, an analogue of the contextualist solution can be given for thought, according to which the premises and conclusion of AI are truly thought, but my true thought that, say, I don’t know that I have hands, had when in the grip of AI, will be compatible with my thought, made in another context, that I do know that very thing. Second, our judgment regarding whether something can or cannot be truly asserted (under appropriate conditions) might be held to affect our judgment regarding whether it’s true or false, even when we make this judgment in solitude, with nothing being said at all. That the premises of AI could be truly asserted, then, makes them (at least) seem true even when they’re just being thought.

My own solution will employ the basic contextualist strategy explained in this section. But, as should be apparent already, we haven’t explained the persuasiveness of AI, and thus haven’t solved our puzzle, if we haven’t located and explained the conversational rule or mechanism by which the skeptic raises (or threatens to raise) the standards for knowledge. And here contextualists have had little to offer.

3 The Subjunctive Conditionals Account (SCA) of the Plausibility of AI’s First Premise

The main stumbling block of other contextualist solutions has been a failure to explain what it is about skeptical hypotheses that makes it so plausible to suppose that we don’t know that they’re false. This point of weakness in the contextualist solutions is the particular point of strength of Nozick’s treatment of AI in his *Philosophical Explanations* (1981). In this and the following

three sections I'll present and defend the *Subjunctive Conditionals Account* (SCA) of the plausibility of AI's first premise, which I've abstracted from Nozick's account of knowledge and skepticism.

According to SCA, the problem with my belief that I'm not a BIV – and I do have such a belief, as do most of us – is that I would have this belief (that I'm not a BIV) even if it were false (even if I were one). It is this that makes it hard to claim to *know* that I'm not a BIV. For, according to SCA, we have a very strong general, though not exceptionless, inclination to think that we don't know that *P* when we think that our belief that *P* is a belief we would hold even if *P* were false. Let's say that S's belief that *P* is *insensitive* if S would believe that *P* if *P* were false. SCA's generalization can then be restated as follows: We tend to judge that S doesn't know that *P* when we think S's belief that *P* is insensitive.

As is well worth noting, this general inclination explains the operation of nonphilosophical skeptical hypotheses that are far less radical than the BIV hypothesis or even the painted mule hypothesis. Just so, it serves to explain why, even though I feel inclined to say that I know the Bulls won their game last night because I read the result in a single newspaper, I still feel strongly pulled toward admitting the (mildly) skeptical claim that I don't know that the paper isn't mistaken about which team won: I realize that my belief that the paper isn't mistaken is a belief I would hold even if it were false (even if the paper were mistaken).

Indeed, after encountering a couple of instances of AI with different skeptical hypotheses plugged into the "*H*" slot (for example, the BIV, the painted mules, and the mistaken paper hypotheses), one develops a sense of what makes for an effective skeptical hypothesis and, thus, an ability to construct convincing instances of AI oneself. To make AI's second premise convincing, it is usually sufficient (though not necessary) that *H* be incompatible with *O*. But what about the first premise? To make *it* convincing, we instinctively look for a hypothesis that elicits in the listener both the belief that the hypothesis doesn't obtain and an acknowledgement that this belief is one she would hold even if the hypothesis *did* obtain.

Upon hearing the hypothesis, typically one can't help but projecting oneself into it. How would things seem to me if that situation obtained? Well, pretty much (or sometimes exactly) as they actually seem to me. And, so, what would I believe

if such a "strange" situation obtained? Pretty much (or exactly) what I actually believe. For example, and in particular, if I *were* a BIV, I would believe every bit as firmly as I actually do that I *wasn't* one. But if this belief is one I would hold even if it were false, how can I be in a position to tell that, or discern that, or *know* that, it's true?

As I've just hinted, a similar explanation, in terms of subjunctive conditionals, can explain the plausibility of the other ways we feel inclined to describe our seemingly limited epistemic position vis-à-vis effective skeptical hypotheses. Consider especially the description involving "ruling out." In a normal zoo setting, most of us would take ourselves to know that the animals in the zebra cage are zebras. From this, it seems, we should be able to infer that they're not cleverly painted mules, since zebras aren't mules. So why are we reluctant to count our seeing the zebras and performing this inference as a case of ruling out the painted mule hypothesis? Because, the explanation goes, even after performing the inference, it still seems we would believe the observed animals weren't painted mules if they were precisely that. Why does it seem we can't tell that they're not painted mules? Because we would believe they weren't even if they were. Ditto for why we seemingly can't discern that they're not and why it seems we can't distinguish their being cleverly painted mules from their not being such, etc.

Also worth noting is the usefulness of SCA in explaining our reluctance to ascribe knowledge in certain lottery situations. Even where the odds of your being a loser are astronomically high (there are 20 million tickets, only one of which is a winner, and you have but one ticket), it can seem that you don't know that you're a loser of a fair lottery if the winner hasn't yet been announced. SCA accounts for this seeming: Your belief that you're a loser is one you would hold even if you were the winner.

SCA is a powerful explanation. But there are problems. As I suggested above, there are exceptions to the general inclination to which SCA appeals: There are cases in which it seems to us that some *S* does know that *P* even though we judge that *S* would believe that *P* even if *P* were false. Some of these exceptions will be quickly discussed in sections 4 and 5 below. The first and main point to make regarding such exceptions, of course, is that this very general inclination needn't be exceptionless to perform the explanatory role SCA assigns it. In section 6 we will see

strong grounds for endorsing SCA as being at least on the right track despite the exceptions to the generalization to which it appeals. But these exceptions are still worth examining, for they will indicate certain important directions in which SCA can be improved, even though we won't be in a position to make SCA ideally precise here.

4 SCA, Grandmothers, and Methods

First, then, consider a case discussed by Nozick:

A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset. Yet this does not mean she doesn't know he is well (or at least ambulatory) when she sees him. (1981, 179)

Here, it seems, the grandmother knows her grandson is well, though it can seem that she doesn't satisfy the third condition of a preliminary form of Nozick's analysis of S knows that *P*, which is:

- (3) If *p* weren't true, S wouldn't believe that *p*.

Nozick's response is to relativize this third condition to the method by which S has come to believe that *P*, yielding:

- (3) If *p* weren't true and S were to use *M* to arrive at a belief whether (or not) *p*, then S wouldn't believe, via *M*, that *p* (179),

where "*M*" is the method by which S has come to believe that *P*.⁹

Unlike Nozick, I'm not presenting an analysis of propositional knowledge. But his grandmother case also seems to be an exception to the general inclination SCA appeals to: Here we're not at all inclined to think the grandmother doesn't know her grandson is well, even though it can seem that if he weren't well, she would still believe he was. The generalization SCA utilizes says that we tend to judge that S doesn't know where S does not satisfy Nozick's third condition for knowledge. One possibility here is to follow Nozick *very* closely by modifying that generalization so that it refers to Nozick's modified, rather than his original, third condition, and thus, like Nozick, explicitly relativizing our account to the method by which S believes that *P*.

Often, though, context takes care of this for us. Even to one aware of the likelihood that the grandmother's family would have kept her in the dark about her grandson's condition were he not well, it *can* seem that even Nozick's initial formulation of the third condition for knowledge is met by the grandmother. On one way of evaluating that simple conditional, it seems that if the grandson were not well, the grandmother would *not* believe he was well. After all, she's looking right at him! The standard possible-worlds semantics for counterfactual conditionals can illuminate what's going on here. When one searches for the possible worlds most similar to the actual world in which the grandson is not well, the respects in which the possible worlds are to resemble the actual world is a highly context-sensitive matter. Especially where the context focuses one's attention on the grandmother and her cognitive and recognitional abilities, one *can* place heavy weight upon similarity with respect to the method she is using to arrive at her belief, and then it can seem that in the closest world in which the grandson is not well, she's looking right at him and seeing that he's not well, and so does *not* believe he is well. On this way of evaluating the conditional, the grandmother *does* satisfy even the initial formulation of Nozick's third condition, and she's no counter-example to the generalization utilized by SCA. But, in evaluating that simple conditional, one can also stress other similarities, particularly ones involving the propensities and plans of the various family members (or whatever facts ground the judgment that if her grandson weren't well, the grandmother would be effectively lied to), to reach the verdict that if he were not well, she *would* believe that he was well.

We can sharpen SCA by specifying that we tend to judge that S doesn't know when she fails to satisfy Nozick's initial formulation of (3), where (3) is evaluated in such a way that heavy emphasis is put upon similarity with respect to the method of belief formation utilized by S, or, following Nozick, we can insert a specification of the method into the antecedent of (3).¹⁰ But in neither case is this to make a very precise modification; rather, it merely indicates the direction in which a more precise account might lie, for any such use of the notion of *methods* of belief formation in our account invites a host of questions (many of which Nozick wrestles with) involving how such methods are to be specified and individuated.

5 SCA and Some Skeptical Hypotheses That Don't Work

Certain instances of AI aren't very persuasive. The first premise of the argument can be quite unconvincing despite the fact that SCA predicts that we'd find it plausible. Suppose, for instance, that in an attempt to show by AI that I don't know I have hands, a skeptic utilizes, instead of the BIV hypothesis, the following simple *H*: I falsely believe that I have hands. The resulting instance of AI seems to pack little or no more punch than a simple skeptic's unsupported claim that I don't know I have hands. It's at the first premise that this ill-fated instance of AI fizzles. But my belief that I don't falsely believe that I have hands is insensitive: If this belief were false (if I did falsely believe that I have hands) I would still believe it was true (I'd still believe that I don't falsely believe that I have hands). Likewise insensitive is my belief that the following hypothesis is false: I'm an intelligent dog who's always incorrectly thinking that I have hands. If this belief of mine were false (if I were such a deluded intelligent dog) I'd still believe it was true (I'd still believe that I wasn't such a creature). So SCA, as it has so far been formulated, predicts that it will seem to us that the above beliefs don't amount to knowledge and that we'll find plausible the first premise of AI that results when the above hypotheses are used. But in fact these instances of AI's first premise are far from convincing. As opposed to the BIV hypothesis, it seems that one *does* know that the deluded dog hypothesis and the simple false belief hypothesis are false.

Again, the main point to make here is that SCA's generalization needn't be exceptionless to be explanatory. While a more precisely Chisholmed refinement of SCA might not have the negations of these ineffective *H*'s as instances of those propositions it says we tend to judge we don't know, I'll here just make a preliminary observation as to what might be going wrong. Part of the problem with these "hypotheses" is that they don't give us much of an idea of *how* I come to have the false belief they assign to me. Hypotheses are supposed to explain; skeptical hypotheses should explain how we might come to believe something despite its being false. The first of these hypotheses simply stipulates that I'm wrong about my having hands, without indicating how I came to be so sadly mistaken. The second adds to the first that I'm a dog, which adds little to

our understanding of how my mistake about having hands came about. By contrast, when we encounter effective skeptical hypotheses, we have some understanding of how (if *H* is true) we have come to falsely believe that *O*. If either of our ineffective hypotheses is filled in so as to make it clear to us how I came to falsely believe I have hands, it becomes effective.

SCA's generalization was this: We tend to judge that *S* doesn't know that *P* when we think that *S*'s belief that *P* is insensitive (when we think that *S* would believe *P* even if *P* were false). The limitation of SCA's generalization that's suggested by these cases is this: We *don't* so judge ourselves ignorant of *P* where not-*P* implies something we take ourselves to know to be false, without providing an explanation of how we came to falsely believe this thing we think we know. Thus, *I falsely believe that I have hands* implies that I don't have hands. Since I do take myself to know that I have hands (*this* belief isn't insensitive), and since the above italicized proposition doesn't explain how I went wrong with respect to my having hands, I'll judge that I do know that proposition to be false. But this again is just a preliminary statement, and there's room for a lot more refinement here. What we need now is some assurance that we're headed in the right direction.

6 SCA Confirmed

Such assurance is to be found by considering what it would take to make it seem to us that we *do* know skeptical hypotheses to be false.

But let's first reconsider the lottery case. As noted above in section 3, we are puzzlingly reluctant to claim knowledge in certain lottery situations. The explanation provided by SCA for this phenomenon is intuitively appealing: It does seem that the fact that we would believe that we were losers even if we were winners is largely what's behind our judgment that we don't know we're losers. SCA receives further powerful support when we consider the grounds that *do* seem to us sufficient for knowledge of one's being a loser. In the lottery situation, even a very minute chance of being wrong seems to deprive one of knowledge. But if we're going to worry about even such minute chances of error, then why does it seem that you do know you're a loser after the winning number has been announced on the radio and

you've compared the numbers on your ticket with the sadly different numbers announced? After all, radio announcements *can* be in error; what you're hearing *may* not be a real radio announcement but the voice of a friend who's rigged up a practical joke; you *might* be suffering from some weird momentary visual illusion and misreading the numbers on your ticket; and so forth. All very remote possibilities, to be sure. But, since we're already countenancing even the most minute chances of error, why don't these possibilities rob us of knowledge even after the announcement has been made and heard?

SCA's explanation of why we don't think we know *before* the announcement is made is that we at that time judge that if we weren't losers, we'd still believe that we were. Note that once you've heard the announcement of the winning numbers and compared them with the numbers on your ticket, it *no longer* seems that if you had been the winner, you'd believe you were a loser. Rather, we judge that in that case you'd now believe you were the winner or would at least be suspending judgment as you frantically double-checked the match. It's very impressive that the very occurrence that would suffice to make it seem to us that you do know you're a loser (the radio announcement) also reverses our judgment regarding the truth of the conditional appealed to in SCA to explain why it seems to us that you don't know before the announcement is made. The occurrence which gets us to judge that we know here also removes what SCA posits as the block to our judging that we know. This is an indication that SCA has correctly identified the block.

SCA similarly provides a very intuitively appealing explanation for why it seems to us that we don't know that skeptical hypotheses are false, as was also noted in section 3. It again receives powerful further confirmation as we look to cases in which one seemingly does know that a skeptical hypothesis doesn't obtain (cases in which skeptical hypotheses that are ordinarily effective fail to be effective). The boastful zoologist I have introduced elsewhere [who invokes his extensive knowledge of zebra and mule anatomy], it seems, knows that the animals in the zebra cage are not cleverly painted mules, while I, it seems, do not. But the very anatomical knowledge that seemingly enables him to know they're not painted mules also has the consequence that if the animals *were* cleverly painted mules, the zoologist, unlike me, would *not* believe that they weren't. And although I

don't seem to know they're not painted mules simply by looking at them, I could, it seems, get to know this if I undertook some special investigation – perhaps, as has been suggested in the literature (Stine 1976, 252), one involving paint remover. *Which* special investigations would do the trick (and under which circumstances would they)? A survey of various scenarios yields an impressive correlation: The investigations that would seemingly allow me to know that the animals aren't painted mules would also affect our judgment as to the truth value of the subjunctive conditional so critical to SCA. Once I have completed the investigation, it seems that I, like the zoologist, would *not* believe that the animals weren't painted mules if in fact they were. Likewise, by checking appropriately independent sources, I could get myself into a position in which I seemingly *would* know that the newspaper isn't mistaken about whether the Bulls won last night. But the checks that would seemingly allow this knowledge would also make it seem that if the paper were mistaken, I would *not* believe it wasn't. Again and again, SCA posits a certain block to our judging that we know, and the changes that would clear the way for our judging that we know also remove this block. This makes it difficult not to believe that SCA is at least roughly correct.

In the case of the BIV hypothesis, it's hard to test SCA in this way, for it's difficult to imagine a situation in which it seems a subject does know that she's not a BIV. But this only confirms SCA: While it's difficult to imagine a situation in which one seems to know that one's not a BIV, it's likewise difficult to imagine circumstances in which the block SCA posits is removed. It's difficult, that is, to imagine a situation in which someone believes they're not a BIV but in which the conditional *If S were a BIV, then S would believe she wasn't a BIV* isn't true. For, as the BIV hypothesis is formulated, one's brain is electrochemically stimulated so that one has precisely those sensory experiences one actually has had. But wouldn't one then have formed precisely those beliefs that one actually has formed, including the belief that one's not a BIV?

It seems that this explanation, SCA, for the plausibility of AI's first premise must be (at least roughly) correct and, therefore, that it points to part of the solution to our puzzle.

Indeed, some readers will wonder why I have claimed only that our general tendency not to

count insensitive beliefs as instances of knowledge explains that premise's plausibility and have stopped short of accepting sensitivity as a necessary condition for knowledge¹¹ and therefore simply endorsing that first premise as true. But while we've just seen strong grounds for simply accepting AI's first premise, there are also strong grounds for accepting AI's second premise and for accepting the denial of its conclusion. We have to stop short somewhere; we can't simply accept all three members of this triad as true. To solve this puzzle, I'll claim that AI's first premise, while not *simply* true, is true according to unusually high standards for knowledge. But, I'll argue, my solution explains why that premise seems true and, more generally, why sensitivity seems necessary for knowledge. If my solution provides the best explanation for how all three members of our puzzling triad seem true, that will be good reason for stopping short where my solution tells us to, rather than where one of its inferior rivals – bold skepticism, for example – tells us to.

7 Nozick's Own Solution and the Abominable Conjunction

Nozick's own treatment of AI, from which SCA was abstracted, fails. This treatment is based on Nozick's account of knowledge as true, *sensitive* belief, where, very roughly, one's true belief that *P* is sensitive to the truth value of *P* if one would not have believed that *P* if *P* had been false.¹² Thus, Nozick's treatment of AI involves accepting the skeptic's first premise. But, at the same time, and much more unfortunately, it also involves denying the second. You *don't* know that you're not a BIV, Nozick claims, because any belief you might have to this effect is insensitive: You would have held this belief even if it were false (even if you were a BIV). By contrast, Nozick claims, your belief that you have hands *is* a sensitive belief: If *it* were false – if you didn't have hands – you would not hold it. So you do know you have hands even though you don't know that you're not a BIV. The skeptic's mistake – the second premise – is supposing that you can know you have hands only if you also know that you're not a BIV.

Or so Nozick claims. This is not the place for a general evaluation of Nozick's analysis of propositional knowledge, so let us confine ourselves to the results of this analysis as applied to the beliefs in

question in AI. Here Nozick's account does very well in issuing the intuitively correct verdict for the relevant particular judgments regarding what is known and what is not. Most of us would judge that we do know such things as that we have hands, and this is Nozick's verdict. And, when a skeptical hypothesis is well chosen, it does seem quite plausible to most of us that we don't know that it doesn't obtain. But there are three relevant issues to our puzzle: Is the first premise of AI true? Is the second premise true? Is the conclusion true? And it's easy to endorse the intuitively correct answer to two out of the three questions if you're willing to take the implausible stand on the remaining one.

Nozick takes his implausible stand on the issue of the second premise, denying it in the face of its evident intuitive appeal.¹³ Accepting his treatment involves embracing the abominable conjunction that while you don't know you're not a bodiless (and handless!) BIV, still, you know you have hands. Thus, while his account does quite well on the relevant particular intuitions regarding what is and isn't known, it yields an intuitively bizarre result on the comparative judgment the second premise embodies.¹⁴

As promised, I won't here rehearse the powerful objections to Nozick's analysis of propositional knowledge that have been put forward,¹⁵ but, assuming that this analysis isn't independently convincing before we turn to the problem of skeptical hypotheses,¹⁶ we're left with little reason to follow Nozick in choosing to take an implausible stand precisely where he has rather than someplace else.

This leaves us in a bind. For, as we saw in sections 3 and 6 above, SCA is quite powerful. That explanation is that we realize that any belief we might have to the effect that an (effective) skeptical hypothesis doesn't obtain is insensitive, and we're inclined to think that insensitive beliefs don't constitute knowledge. How can we appropriate that explanation without following Nozick in having to implausibly deny the second premise of AI and embrace the abominable conjunction?

8 Strength of Epistemic Position and AI's Second Premise

Here's how: by incorporating SCA into a contextualist solution to our puzzle that avoids such a

fumbling of AI's second premise. Indeed, I propose a *very* strong endorsement of that second premise.

Recall that according to contextualist theories of knowledge attributions, how strong a subject's epistemic position must be to make true a speaker's attribution of knowledge to that subject is a flexible matter that can vary according to features of the speaker's conversational context. Central to contextualism, then, is the notion of (*relative*) *strength of epistemic position*. In presenting and defending contextualism, I've found that most listeners feel that they understand pretty well what's meant when I claim, for instance, that sometimes the standards for knowledge are higher than usual, or that in some conversational situations one's epistemic position must be stronger than in others to count as knowing. But it would be good to clarify this important notion of strength of epistemic position as best we can by, for instance, supplying an intuitive test for when one epistemic position is stronger than another. The best such device is that of *comparative conditionals*. One can have a variety of grounds for assenting to conditionals like *If Mugsy is tall, then Wilt is tall*, and *If Wilt is not tall, then Mugsy is not tall*. But one very good basis for assenting to these conditionals is the comparative knowledge that Wilt is at least as tall as Mugsy. Likewise, where *S* is a putative subject of knowledge, *P* is a true proposition that *S* believes, and *A* and *B* are situations in which *S* is found, we can have similarly comparative grounds for assenting to conditionals of the form *If S knows that P in A, then S knows that P in B*. In such a case, the comparative grounds for our assent is our realization that *S* is in *at least as strong* an epistemic position with respect to *P* in situation *B* as he is in with respect to that same proposition in situation *A*, and this comparative conditional serves as a good intuitive test for that comparative fact: It brings that fact to light.

So, for instance, to borrow some examples from Alvin Goldman (1976), let Henry be our subject, and let *What Henry is seeing is a barn* be the thing Henry putatively knows. Both in situation *F* (for "fakes") and in situation *N* ("no fakes"), Henry is driving through the countryside and, having no reason to think there's anything unusual going on, very firmly believes, and takes himself to know, that the object he's seeing is a barn. And indeed, in both cases, it is a barn. But in *F*, unbeknownst to him, Henry is in an area that is filled with very convincing fake barns – papier-

mâché barn façades. In fact, we may suppose that Henry has just been fooled more than twenty times by such fakes, although he's now looking at the only actual barn for miles around, and so this time truly believes that what he's seeing is a barn. *N* is exactly like *F*, except that there are no fakes in the area – the things Henry has taken to be barns have all actually been barns. With regard to these examples, the conditional *If Henry knows in F, then he knows in N* seems to get the comparison right, indicating that Henry's in at least as strong an epistemic position in situation *N* as he is in situation *F*. The evident failure of *If Henry knows in N, then he knows in F* to get the comparison right shows that Henry's not in as strong a position to know in *F* as in *N*. Together, these results indicate that Henry's in a stronger epistemic position in *N* than in *F*.

As is important to our discussion of AI's second premise, comparative conditionals can similarly be used to test the relative strength of epistemic position of a single subject with respect to *different propositions* that subject believes in the same situation: Thus, the intuitive correctness of *If S knows that P, then S knows that Q* and *If S doesn't know that Q then S doesn't know that P* can indicate that *S* is in at least as strong an epistemic position with respect to *Q* as she's in with respect to *P*.¹⁷

Sometimes no clear verdict results when we attempt to evaluate a conditional in this comparative way, for the good reason that it's unclear how the two epistemic positions we're evaluating compare with one another. Thus, if we compare a situation in which Henry has a good look at the barn but in which there are a couple of fake barns several miles away that Henry hasn't encountered with a situation in which there are no fakes at all in Henry's vicinity but in which he doesn't have quite as good a look at the barn, the relevant conditionals can be difficult to evaluate. But, in many instances, some of the relevant conditionals are clearly true on comparative grounds.

Such is the case with instances of AI's second premise, where the skeptical hypothesis is well chosen. They seem true and are true, I suggest, for just this comparative reason: As we realize, we are in at least as good a position to know that the hypothesis is false as we're in to know the targeted piece of presumed ordinary knowledge.¹⁸ Let's look briefly at some instances. Recall the following epistemologically perplexing pairs of propositions:

<i>not-H</i>	<i>O</i>
I'm not a BIV.	I have hands.
Those animals aren't just cleverly painted mules.	Those animals are zebras.
The paper isn't mistaken about whether the Bulls won last night.	The Bulls won last night.

Given natural background assumptions, we can sense that the following comparative fact holds for each of the above pairs: I am in no better a position to know that *O* than I am in to know that *not-H*. This comparative fact is revealed in each case by the highly plausible conditional that is AI's second premise: If I don't know that *not-H*, then I don't know that *O*. Closely tied to that comparative fact in each case is the related and intuitively compelling realization that it would be no wiser to bet one's immortal soul on *O*'s being true than to bet it on *not-H*'s being true.

I propose then to accept the relevant conditional with respect to each of the above pairs, and to accept other convincing instances of AI's second premise. Indeed, these conditionals are true *regardless of how high or low the standards for knowledge are set*. Just as the comparative fact that Wilt is at least as tall as Mugsy has the result that the conditional *If Wilt is not tall, then Mugsy is not tall* will be true regardless of how high or low the standards for tallness are set so the comparative fact that I'm in at least as strong an epistemic position with respect to *not-H* as I'm in with respect to *O* will result in *If I don't know that not-H, then I don't know that O* being true regardless of how high or low the standards for knowledge are set. Thus, we will never have to follow Nozick in accepting the abominable conjunction: that conjunction is false at any epistemic standard.

With that ringing endorsement of AI's second premise anchored firmly in place, we can return to the first premise, hoping to incorporate SCA into a contextualist account of that premise's plausibility.

9 Strength and Sensitivity

As has become very apparent, two notions that are central to my attempt to solve our puzzle are, on the one hand, the Nozickean notion of the sensitivity of beliefs and, on the other, the notion of strength of epistemic position. While both notions stand in need of a good deal of sharpening and explanation (only some of which they'll receive

here), we've already obtained interesting results applying them to the epistemologically perplexing pairs of propositions displayed above. In each case, one's belief in *O* is sensitive, while one's belief in *not-H* is insensitive. Yet, at the same time, one is in at least as strong an epistemic position with respect to *not-H* as one is in with respect to *O*.

For each of the second and third pairs of propositions, one could gather further evidence, strengthen one's epistemic position with respect to both *not-H* and *O*, and *make* even one's belief that *not-H* sensitive. But even before this further evidence is gathered, one's belief that *O* is *already* sensitive, despite the fact that one is in no stronger an epistemic position with respect to this *O* than one is in with respect to *not-H*. (With respect to the first pair of propositions, it is difficult to imagine a situation in which one is in such a strong position with respect to one's not being a BIV that this belief is sensitive.)

This leads us to an important insight regarding skeptical hypotheses: One's epistemic position with respect to propositions to the effect that skeptical hypotheses don't hold must be stronger than it is with respect to other, more ordinary propositions (e.g., our above *O*s) if belief in such propositions is to be sensitive.

An explanation of our two central notions in terms of possible worlds will provide a partial and quite rough-and-ready, but still somewhat enlightening, picture of how this situation can arise. An important component of being in a strong epistemic position with respect to *P* is to have one's belief as to whether *P* is true match the fact of the matter as to whether *P* is true, not only in the actual world, but also at the worlds sufficiently close to the actual world. That is, one's belief should not only be true, but should be non-accidentally true, where this requires one's belief as to whether *P* is true to match the fact of the matter at nearby worlds. The further away one can get from the actual world, while still having it be the case that one's belief matches the fact at worlds that far away and closer, the stronger a position one is in with respect to *P*. (Recalling the results of section 4, we should remember either to restrict our attention solely to those worlds in which the subject uses the same method of belief-formation she uses in the actual world, or to weigh similarity with respect to the subject's method very heavily in determining the closeness of possible worlds to the actual world.) If the truth-tracking of one's belief as to whether *P* extends far enough from

actuality to reach the closest not- P worlds, then one doesn't believe that P in those closest not- P worlds, and one's belief that P is sensitive. But how far from actuality must truth-tracking reach – how strong an epistemic position must one be in – to make one's belief that P sensitive? That, of course, depends on how distant from actuality the closest not- P worlds are.

Consider my belief that I have hands. I believe this at the actual world, and it's true. What's more, in the other nearby worlds in which I have hands, I believe that I do. There are also, at least in my own case, some alarmingly close worlds in which I don't have hands. These include worlds in which I lost my hands years ago while working on my uncle's garbage truck. In the closest of these not- P worlds, I'm now fully aware of the fact that I'm handless, and my belief as to whether I have hands matches the fact of the matter. My belief as to whether I have hands doesn't match the fact in various worlds in which I'm a BIV, of course, but these are *very* distant. While there are closer worlds in which the match fails, it seems that in a fairly wide range of worlds surrounding the actual world, my belief as to whether I have hands does a good job of matching the fact of the matter. Thus, I'm in a pretty strong epistemic position with respect to that matter.

Now let P be *I'm not a BIV*. Where not- P (here, *I am a BIV*) is quite remote, one can be in a quite strong epistemic position with respect to P merely by believing that P in all the nearby worlds. As I do believe this P in such nearby worlds, I'm in a pretty strong epistemic position with respect to this P . This can occur, and in my case, does occur, even though one's belief as to whether P doesn't match the fact of the matter in the closest not- P worlds: Since even the closest of the not- P worlds are quite distant, one's belief as to whether P needn't match the fact of the matter that far from the actual world for one to be in a quite strong position with respect to P .

But for one's belief that P to be sensitive, one must *not* believe that P in the closest not- P worlds. Since skeptical hypotheses tend to fasten on somewhat remote (and sometimes very remote) possibilities, then, one can be in a relatively (and sometimes a very) strong position with respect to beliefs to the effect that they don't obtain (since one's belief as to whether they obtain matches the fact of the matter over a wide range of worlds closest to the actual world), while these beliefs remain insensitive (since one would still believe

that the hypotheses didn't obtain in the closest worlds in which they do obtain). By contrast, where P is such that there are both P and not- P worlds very close to the actual world, one's belief that P must be sensitive (one must not believe that P in the closest not- P worlds) in order for one to be in even a minimally strong epistemic position with respect to P , and, conversely, one needn't be in a very strong position for one's belief to be sensitive.

10 The Rule of Sensitivity and the Beginnings of a New Contextualist Solution

The important insight regarding skeptical hypotheses – that one's epistemic position with respect to propositions to the effect that skeptical hypotheses don't hold must be stronger than it is with respect to other propositions before beliefs in such propositions can be sensitive – suggests a new contextualist account of how, in presenting AI, the skeptic raises the standards for knowledge. Let's call the conversational rule this new account posits as the mechanism by which the skeptic raises the standards for knowledge the "Rule of Sensitivity." Although a more general formulation of this rule is desirable, I will here state it in such a way that it applies only to attributions (and denials) of knowledge, since such applications are what's needed to address the present puzzle.¹⁹ So limited, our rule is simply this: When it is asserted that some subject S knows (or does not know) some proposition P , the standards for knowledge (the standards for how good an epistemic position one must be in to count as knowing) tend to be raised, if need be, to such a level as to require S 's belief in that particular P to be sensitive for it to count as knowledge. Where the P involved is to the effect that a skeptical hypothesis does not obtain, then this rule dictates that the standards will be raised to a quite high level, for, as we've seen, one must be in a stronger epistemic position with respect to a proposition stating that a skeptical hypothesis is false – relative to other, more ordinary, propositions – before a belief in such a proposition can be sensitive.

A story in terms of possible worlds again provides a rough-and-ready, but still perhaps enlightening, picture of how the Rule of Sensitivity operates. Context, I've said, determines how strong an epistemic position one must be in to

count as knowing. Picture this requirement as a contextually determined sphere of possible worlds, centered on the actual world, within which a subject's belief as to whether *P* is true must match the fact of the matter in order for the subject to count as knowing. (Given the results of section 4, we must again remember either to restrict our attention solely to those worlds in which the subject uses the same method of belief formation she uses in the actual world, or to weigh similarity with respect to the subject's method very heavily in determining the closeness of possible worlds to the actual world.) Call this sphere the sphere of epistemically relevant worlds. As the standards for knowledge go up, the sphere of epistemically relevant worlds becomes larger – the truth-tracking of one's belief must extend further from actuality for one to count as knowing. Given this picture, the Rule of Sensitivity can be formulated as follows: When it's asserted that *S* knows (or doesn't know) that *P*, then, if necessary, enlarge the sphere of epistemically relevant worlds so that it at least includes the closest worlds in which *P* is false.

A powerful solution to our puzzle results when we follow the basic contextualist strategy (see section 2) and utilize this Rule of Sensitivity to explain how the standards for knowledge are raised by the skeptic's presentation of AI. While many noteworthy features and virtues of this solution are best explained by comparing it with the other proposed solutions to our puzzle, as I'll do in following sections, the basic idea of the present solution is this. In utilizing AI to attack our putative knowledge of *O*, the skeptic instinctively chooses her skeptical hypothesis, *H*, so that it will have these two features: (1) We will be in at least as strong a position to know that not-*H* as we're in to know that *O*, but (2) Any belief we might have to the effect that not-*H* will be an insensitive belief (a belief we would hold even if not-*H* were false – that is, even if *H* were true). Given feature (2), the skeptic's assertion that we don't know that not-*H*, by the Rule of Sensitivity, drives the standards for knowledge up to such a point as to make that assertion true. By the Rule of Sensitivity, recall, the standards for knowledge are raised to such a level as to require our belief that not-*H* to be sensitive before it can count as knowledge. Since our belief that not-*H* isn't sensitive (feature (2)), the standards are driven up to such a level that we don't count as knowing that not-*H*. And since we're in no stronger an epistemic position with respect to *O* than we're in with respect to

not-*H* (feature (1)), then, at the high standards put in place by the skeptic's assertion of AI's first premise, we also fail to know that *O*. At these high standards, the skeptic truthfully asserts her second premise (which, recall, is also true at lower standards), and then truthfully asserts AI's conclusion that we don't know that *O*.²⁰ This accounts for the persuasiveness of AI. But since, on this account, the skeptic gets to truthfully state her conclusion only by raising the standards for knowledge, AI doesn't threaten the truth of our ordinary claims to know the very *O*s our knowledge of which the skeptic attacks. For the fact that the skeptic can install very high standards that we don't live up to has no tendency to show that we don't satisfy the more relaxed standards that are in place in more ordinary conversations and debates.

11 Our New Contextualist Solution Clarified and Compared with the Straightforward Solution

The puzzle of skeptical hypotheses, recall, concerns the premises of AI together with the negation of its conclusion:

1. I don't know that not-*H*.
 2. If I don't know that not-*H*, then I don't know that *O*.
- not-*C*. I do know that *O*.

A solution to the puzzle must, of course, issue a verdict as to the truth of each of these three, but it must also explain why we find all of them plausible.

Let's be clear about what our present contextualist solution has to say about each of these. Our verdict regarding (2) is that it's true regardless of what epistemic standard it's evaluated at, so its plausibility is easily accounted for. But this, combined with a similarly enthusiastic endorsement of (1), would land us in bold skepticism. We avoid that fate by endorsing (1) as true, not at all standards, but only at the unusually inflated standards conducive to skepticism. Thus, on our solution, we do know, for instance, that we're not BIVs, according to ordinary low standards for knowledge. But, though (1) is false when evaluated according to those ordinary low standards, we're able to explain its plausibility, as we've seen, by means of the fact that the high standards at which (1) is true are precisely the standards that an assertion or denial

of it put into play. Since attempts to assert (1) are bound to result in truth, and attempts to deny it are destined to produce falsehood,⁽²¹⁾ it's no surprise that we find it so plausible.

But what of (not-C)? On the present solution, claims to know ordinary propositions are true according to ordinary low standards but false according to the highly inflated standards that, by the Rule of Sensitivity, are put in place by the assertion of (1). (Not-C) seems plausible because it's true when evaluated at the standards most normally applied to it. But, it will be asked, why do we find these claims to know plausible even when we're in a context in which the skeptic has raised the standards to such a level that these claims are false? A little caution is in order here. It's controversial just how intuitively correct (not-C) does seem to us in such a context. Most of us feel some ambivalence. Such ambivalence is to be expected whenever we're dealing with a puzzle consisting of mutually inconsistent propositions, all of which are individually plausible. For when the propositions are considered together, one will have this good reason for doubting each of them: that the others seem true. And it's difficult to distinguish the doubt of (not-C) that arises from this very general source (that its falsehood follows from other things one finds plausible) from that which arises from the fact that the standards are high. At any rate, the very strong pull that (not-C) continues to exert on (at least most of) us even when the standards are high is explained in the manner outlined in section 2: Even while we're in a context governed by high standards at which we don't count as knowing that *O*, we at the same time realize that as soon as we find ourselves in more ordinary conversational contexts, it will not only be true for us to claim to know these very *O*s that the skeptic now denies we know, but it will also be wrong for us to deny that we know these things. It's easy, then, to think that the skeptic's present denial must be equally false and that it would be equally true for us now, in the skeptic's presence, to claim to know that *O*.

The verdicts the present solution issues regarding the truth values of the members of the triad are complicated by the fact that ours is a contextualist solution. Only (2) receives the same verdict regardless of what the epistemic standards are; the truth values of (1) and (not-C) vary with context. It's just this variance that our solution so essentially relies on in explaining how we fall into our puzzling conflict of intuitions. Noncontextu-

alist (henceforth, "straightforward") solutions, on the other hand, must choose one of the members of this triad to deny, claiming this loser to be false according to the invariant epistemic standards that govern all attributions and denials of knowledge: The "Moorean" solution in this way denies (1),²² the "Nozickean" (2), and the "Bold Skeptical" solution thus denies (not-C), accepting that we speak falsely whenever, even in ordinary, nonphilosophical discussions, we claim to know the *O* in question.

From the perspective of our present contextualist solution, each of these straightforward solutions results in part, of course, from a failure to see the truth of contextualism. But which straightforward solution an invariantist confusedly adopts will depend on the standards that dominate her evaluation of our beliefs in *O* and in not-*H*. If her evaluation is dominated by the relatively low standards that govern our ordinary, out-on-the-street talk of knowledge, she will end up a Moorean. If she evaluates the beliefs in question according to the high standards that are put into place by the skeptic's presentation of AI, bold skepticism is the result. The Nozickean solution ensues from evaluating each belief according to the standards that would most often be used in evaluating that belief. For reasons we've seen, a claim to know (or an admission that one doesn't know) that a skeptical hypothesis is false will, by the Rule of Sensitivity, tend to invite a very high reading, at which the admission is true and the claim is false. But a claim to know that *O* doesn't so demand a high reading. From the present perspective, the Nozickean is reacting to the fact that one can usually truthfully claim that one does know that *O* and can usually truthfully claim not to know that not-*H*. What the Nozickean misses is how difficult it is to make these two claims together: once you have admitted that you don't know that not-*H*, it seems the reverse of intuitively correct to claim to know that *O*, at least until the conversational air is cleared.

To succeed, a straightforward solution must explain what leads our intuitions astray with respect to the unlucky member of the triad which that solution denies. Otherwise, we'll have little reason for denying just that member of the triad. Nozick himself provides no such explanation with respect to (2), parenthetically leaving this vital task to "further exploration,"²³ and other Nozickeans, if any there be, have not, to the best of my knowledge, progressed any farther along this front.

Mooreans, to the best of my knowledge, have fared no better in explaining why we're so reluctant to claim the status of knowledge for our insensitive beliefs. It's the defenders of bold skepticism who've made the most progress here. In the remaining sections, I'll explain why our contextualist solution is superior to that of the bold skeptic.

12 Bold Skepticism and the Warranted Assertability Maneuver

Almost of all of the time, it seems to almost all of us that we do know the *O*s that the skeptic claims we don't know. According to the bold skeptic, whenever we say or think that we know these things, we say or think something false. The bold skeptic thus implicates us, speakers of English, in systematic and widespread falsehood in our use, in speech and in thought, of our very common word "know." Equally paradoxically, the bold skeptic holds that we're speaking the truth whenever we say that someone doesn't know these *O*s, even though it seems to most of us that we'd then be saying something quite false. What leads us astray? Peter Unger and Barry Stroud have suggested on behalf of bold skepticism that although we don't know these *O*'s, it's often useful for us to claim that we do know them, and we are therefore often warranted or justified in making such claims. What then leads us astray is this: We mistake this useful/justified/warranted assertability of knowledge ascriptions for truth.²⁴ On the other side of the coin, presumably, we're mistaking the useless/unwarranted/unjustified assertability of denials of knowledge for falsehood.

Two serious problems emerge for the bold skeptic at this point. The first is that such "warranted assertability maneuvers" could be attempted by advocates of the other solutions as well. Warranted assertability indeed can be mistaken for truth, and unwarranted assertability for falsehood, but this by itself does not favor the bold skeptic's solution over the other straightforward approaches. Each of the straightforward approaches denies a member of the triad constituting our puzzle, and each it seems could claim that the reason this loser they've chosen seems true, though it's in fact false, is that we're often warranted in asserting it, and we mistake this warranted assertability for truth. Thus, the Moorean, for instance, could claim that although we do indeed know that *H* is false, we're not warranted in claiming that we know

this (though this claim would be true), but are rather warranted in saying that we don't know (though this latter is false). Simply attributing apparent truth to warranted assertability is a game almost any party to this dispute can fairly easily play.²⁵ That this line of thought would eventually work out any better for the bold skeptic than for his opponents would take some showing.²⁶

It's at (1) that the skeptic has his best hope of gaining an advantage over my solution, for that premise indeed does seem true, and, unlike the skeptic, I've stopped short of fully endorsing it, making do with an explanation of its plausibility. But the skeptic's other problem lurks here. Usually, while solving a philosophical puzzle consisting of a set of individually plausible but mutually inconsistent claims, one only has to explain (away) the plausibility of those members of the set one denies, and one is relieved of the burden of explaining the plausibility of those members that one endorses, their truth and our ability to recognize that truth being explanation enough of their apparent truth. But truth does not suffice to explain apparent truth where one makes us out to be absolutely horrible judges of truths of the kind in question. Thus, the skeptic's second big problem is that, because he holds that we're subject to constant and radical error as to the scope of our knowledge, consistently thinking we know things when we don't, the skeptic, although he thinks (1) is true, owes us an explanation for its plausibility. Given that our habit of mistaking our ignorance for knowledge is so pervasive, why doesn't it seem to us *here* that we know what, in fact, we don't – that these skeptical hypotheses are false? Why does our lack of knowledge, which we're so pervasively blind to, shine through so clearly to us just where the issue is whether we know a skeptical hypothesis to be false?

The skeptic's initial answer will certainly be that we're *not* warranted in claiming to know that skeptical hypotheses don't obtain, and thus can't mistake warranted assertability for truth here. But then, to see why skeptical hypotheses are effective, we must be told why we're not warranted in claiming to know that skeptical hypotheses are false, given that, according to the skeptic, we are warranted in claiming to know all manner of other things that in fact we don't know. And here skeptics have little to offer. But if the results of sections 3 and 6 above are correct, the answer must involve the lack of sensitivity enjoyed by our beliefs that

skeptical hypotheses don't obtain. The skeptic's use of SCA will take this form: Although we know nothing (or very little), it's when our beliefs are insensitive that we're not even warranted in asserting that we know and we therefore recognize our lack of knowledge. But the skeptic must now also address AI's second premise, making sure his endorsement of SCA is made in such a way as to account for our intuitions here. Indeed, whether or not he buys into SCA, the skeptic faces this question: If, as he claims, we're usually under the delusion that we know that *O*, but we customarily recognize that we don't know that not-*H*, why aren't we happy to conjoin this error with that insight and embrace the abominable conjunction?

This may look like a difficult question, but the skeptic has a ready answer. His problem is that the warranted assertability maneuver by itself didn't really solve our puzzle, but rather re-introduced it in a new form. And the only way I've seen to incorporate SCA into a treatment of AI that also handles the other pieces of our puzzle is to employ the idea that contextually sensitive epistemic standards govern our use of "know," and to posit the Rule of Sensitivity as the mechanism by which the AI skeptic drives those standards up, as I've advocated here. But wise invariantists typically accept that contextually varying standards govern our use of ascriptions and denials of knowledge. The sensible invariantist will admit that, of course, what passes for knowledge in some contexts won't so pass in others. Being an invariantist, he'll deny that the truth conditions of knowledge attributions vary in the way the contextualist claims they do. But the clever invariantist will maintain that the varying epistemic standards that the contextualist supposes govern the truth conditions of these sentences in fact govern their conditions of warranted assertability.²⁷

This allows the bold skeptic to mimic any contextualist solution, and in particular the solution I'm advocating here, by means of a simple twist. With respect to my solution, the bold skeptic can maintain that the Rule of Sensitivity is a rule for the raising of the epistemic standards governing our use of sentences ascribing knowledge to subjects, alright, but insist that it governs the warranted assertability conditions of these sentences, rather than their truth conditions, which, he'll maintain, remain constant at a level beyond the reach of mere mortals to satisfy. The warranted assertability maneuver can then be employed: We mistake warranted assertability for truth (and

unwarranted assertability for falsehood). Thus, since we're never warranted in claiming to know that skeptical hypotheses don't obtain (due to the operation of the twisted Rule of Sensitivity), we're led to judge (correctly) that such claims to knowledge would be false. And since AI's second premise is always warranted, we judge (again correctly) that this premise is true. But since a claim to-know some *O* is usually warranted, due to the low standards for warranted assertability that would ordinarily be applied to such a claim, we judge (incorrectly) that we know this *O*. Thus, my solution, like other contextualist solutions, can be easily adapted to suit the purposes of the bold skeptic. The result is a theory parallel to my own contextualist solution, which differs in its semantics of "know": According to this parallel invariantist theory, the context-sensitive varying epistemic standards we've discovered govern the warranted assertability conditions of attributions and denials of knowledge, rather than their truth conditions, which are held to be invariant.²⁸ How shall we rationally decide between a contextualist solution, and in particular the one I'm here defending, and the bold skeptic's analogue of it?²⁹

13 Bold Skepticism and Systematic Falsehood

Like its contextualist relatives, our new solution is designed largely with the goal in mind of crediting most of our attributions of knowledge with truth. And no wonder. We in general take it as a strike against a theory of a common term of a natural language that it involves the speakers of that language in systematic and widespread falsehood in their use of that term. Let's borrow an example and suppose, for instance, that a crazed philosopher claimed that there are no physicians, because, in addition to holding a medical degree, a necessary condition for being a physician is that one be able to cure any conceivable illness.³⁰ On what grounds should we reject this bizarre conjecture in favor of a more traditional and less demanding account of what it is to be a physician? Our language certainly could have been such that *S*'s having the ability to cure any conceivable illness was a truth condition of "*S* is a physician" (although the word "physician" would not have been very useful in that case). In virtue of what is our language in fact such that the strange theory is not true of it? I'm of course not in a position to

give a complete answer to this question, but it's eminently reasonable to suppose that such facts as these, regarding our use, in thought and in speech, of the term "physician" are involved: that we take to be physicians many licensed practitioners of medicine who don't satisfy the demanding requirement alleged; that we seriously describe these people as being physicians; that we *don't deny* that these people are physicians; etc. It's no doubt largely in virtue of such facts as these that the traditional view, rather than the conjecture of our crazed philosopher, is true of our language. (The correctness of the traditional view largely *consists* in such facts.) And these facts also provide us with our best reasons or evidence for accepting the traditional, rather than the strange, hypothesis regarding the semantics of "physician." In this case, that the peculiar theory implicates us in systematic and widespread falsehood in our speech and thought involving "physicians" is a (constitutive and evidential) strike against the theory that proves quite decisive.

If our crazed philosopher tried to account for the above facts regarding our use of the term "physician" via the quick and easy conjecture that the less demanding requirements that are more traditionally assigned to "physician," while they don't accurately specify the truth conditions of sentences involving that term, do articulate these sentences' warranted assertability conditions, we should not, on the basis of this maneuver, suspend our judgment against his contention. That his theory involves us in systematic falsehood continues to constitute a strike against it, and in the absence of quite weighty counterbalancing considerations that favor the strange theory over the traditional one, this strike remains decisive.

Of course, the problem with this hopeless non-starter of a theory is that there don't seem to be any such counterbalancing considerations in its favor. By contrast, bold skepticism can appear to be supported by skeptical arguments like AI. Though the bold skeptic's resolution of our puzzle involves us in systematic falsehood because of its unwavering acceptance of AI's conclusion, it at the same time can seem to make sense of *other* pieces of the puzzle (that we're inclined to say that we don't know that skeptical hypotheses are false and to say that we don't know various ordinary things if we don't know these hypotheses to be false), making the warranted assertability maneuver seem more motivated here than it is in the hands of our imagined crazed philosopher. But, as we

saw in the previous section, this appearance is deceptive. Bold skepticism, by itself, does not explain the plausibility of AI's premises. To help the skeptic solve the puzzle, I've had to ascribe to him an analogue of our new solution.³¹ But once we see that the skeptical puzzle can be solved just as well without the bold skeptic's systematic falsehood, we're left with no reason for paying that high price for a solution.³² Indeed, since the bold skeptical solution and our new contextualist solution under consideration closely parallel each other, there's not much difference in how they solve the puzzle. That the bold skeptical resolution involves us in systematic falsehood is one of the few differences to be found here, and it's a weighty consideration against that resolution. And, with there being little room for weighty compensating advantages for this resolution over the contextualist's (given how similar they are in other respects), this consideration proves decisive. So, as with the crazed philosopher's theory of "physician," the bold skeptic's resolution of AI should be rejected because it involves us in systematic and widespread falsehood in our use of a common term of our language.

14 Begging the Question Against the Skeptic?

If skeptics are allowed to play King of the Mountain – they start off on top (never mind how they got there) and it's the anti-skeptics' job to knock them off – displacing them can be a very difficult task. How difficult depends on several factors, one of which is what premises the anti-skeptic is allowed to appeal to in an argument designed to dethrone the skeptic. If the skeptic won't allow any premises to be available, then, as Thomas Reid noted, "It would be impossible by argument to beat him out of this stronghold; and he must even be left to enjoy his scepticism" (1895, 447).³³ If, to make the game a bit more interesting, a slim range of claims is allowed to pass inspection and be available for use in the anti-skeptical campaign, then (as Reid again recognized) it's often difficult to say what, if anything, of importance would follow from the fact that the skeptic can or cannot be knocked from his perch by arguments from premises of that particular type.

I have little interest in playing King of the Mountain. But skeptical arguments like AI threaten to show that the skeptic needn't just play this

game, but can *gain* the top of the mountain – that starting from our own beliefs and intuitions, he can give us better reasons for accepting his skepticism than we have for rejecting it. I've here argued that the bold skeptic cannot win *this* battle – that of providing the best resolution of our puzzling conflict of intuitions. Although AI's premises are initially plausible, the best resolution for the conflict of intuitions generated by AI is not that of the bold skeptic.

Along the way, I've been assuming certain things that we believe but that the skeptic claims we can't know, thereby perhaps raising the concern that I'm begging the question against the skeptic. For instance, in claiming that my belief that I have hands is sensitive, I betray my conviction that I'm not a BIV, either in the actual world or in any nearby worlds. Indeed, I'm ready to admit to the skeptic that if I am a BIV, then I don't know I have hands, according to any standards for knowledge. But, of course, as I firmly believe, I'm not a BIV.

Notes

- 1 I choose this *O* partly for its historical connections to Descartes's First Meditation, and also because I think it is an exemplary case of something we ordinarily think we know. But while we would ordinarily think we know this *O*, we'd seldom have occasion to *say* that we know it, because cases in which such a claim to knowledge would be conversationally in order are quite rare. (Exception: A teacher begins an epistemology lecture by matter-of-factly listing various things she knows, and that any plausible theory of knowledge should make her come out to know. In the course of this listing, she says, "And I know that I have hands.") For this and various related reasons, some might not like my choice of *O*. Such readers are invited to supply their own favorite exemplary cases of things we know as the skeptic's target.
- 2 Those who think that Hilary Putnam may have already disarmed BIV-inspired skepticism should understand the BIV hypothesis to be the hypothesis that one's brain has been *recently* envatted after many years of normal embodiment. For even if Putnam is right in claiming that the content of the beliefs of the BIVs of his scenario is such that these BIVs aren't massively deceived, it seems that recently envatted BIVs are so deceived.
- 3 AI takes its name primarily from its first premise. But since one of AI's best formulations (to which I hereby refer readers seeking a good version of AI that has not been so brutally pared) is in chapter 1 of Peter

Is it legitimate for me to use this conviction in a debate against the skeptic? Not if we're playing King of the Mountain. But if the skeptic is marshalling deeply felt intuitions of ours in an attempt to give us good reasons for accepting his skepticism, it's legitimate to point out that other of our beliefs militate against his position, and ask why we should give credence to just those that favor him. And if we can further show that those beliefs that seem to favor his solution can be accommodated in our solution better than he can accommodate those of our beliefs that are hostile to him, the best conclusion we can draw is that we're *not* ordinarily mistaken when we claim or ascribe knowledge, despite the bold skeptic's attempt to show that we are. Instead, the main insights to be drawn from a study of AI involve the context-sensitivity of attributions of knowledge, and the role that the Rule of Sensitivity plays in changing the epistemic standards that govern these attributions.

Unger's book *Ignorance: A Case for Skepticism* (1975), it is in more than one sense that it is an argument "from ignorance."

- 4 I actually haven't pared AI to its *barest* essentials. It could be further pared to a one-premise argument: I don't know that not-*H*; so, I don't know that *O*. The second, "bridge" premise has been added to facilitate my treatment of the argument, nicely dividing those issues that impact on the acceptability of the first premise from those germane to the second.

AI is the first and great argument by skeptical hypothesis. And the second, like unto it, is *The Argument from Possibility* (AP), which like AI, takes its name from its first premise, and which has this form:

1. It is possible that H_{ind} .
 2. If it is possible that H_{ind} , then it is possible that not- O_{ind} . So,
 3. It is possible that not- O_{ind} .
 4. If it is possible that not- O_{ind} , then I don't know that *O*. So,
- C. I don't know that *O*.

(The subscript "ind" indicates that what occurs in the scope of "It is possible that" is to be kept in the indicative mood, so that the possibility expressed will be an epistemic one. The "bridge" premises, 2 and 4, can be omitted.) In this paper I address only AI, but let me quickly indicate how AP should be handled.

Premise 4, which initially strikes many as AP's weakest link, is actually correct (DeRose 1991, section G). Thus, the AP skeptic must be stopped *before* she reaches step 3. Fortunately, the treatment of AI that I present in this paper can be generalized to handle the initial phase (steps 1–3) of AP as well. This treatment of AP is left here as an exercise for the reader, but is explained in chapter 3, especially section K, of my 1990.

- 5 This is especially true of Stewart Cohen, to whom I'm indebted for his general setup of the puzzle as a conflict of intuitions, a satisfactory solution of which requires an explanation of why the puzzle arises. See Cohen 1988, pp. 93–4.
- 6 For a bit more on the nature of contextualist theories, see my 1992. The notion of (comparative) strength of epistemic position, central to my characterization of contextualism, will be explicated below in sections 8 and 9.
For exemplary contextualist treatments of the problem of skepticism, see especially Unger 1986 and Cohen 1988.
- 7 This is at least so according to *skeptic-friendly* versions of contextualist solutions, as will be explained later in this section.
- 8 Thanks to Richard Grandy and to Peter Unger for pressing this point.
- 9 Precisely, what Nozick does is this: He analyzes the technical locution "S knows, via method *M*, that *p*," and then in turn analyzes the relation of S's knowing that *p* in terms of this technical locution. The revised third condition I've displayed is part of Nozick's attempt to analyze the technical locution.
- 10 These are not identical modifications. On the first option, similarity with respect to method is weighted heavily, but can be outweighed by other factors. Thus, even so evaluated, the most similar world(s) in which the antecedent of the original (3) are true may be worlds that diverge from the actual world with respect to the method by which S came to believe that *P*. By contrast, on the second option, since the method by which S believes that *P* becomes part of the antecedent of the conditional we're evaluating (the modified (3)), the closest possible world(s) in which that antecedent is true cannot be worlds that diverge from the actual world with respect to method.
- 11 Or, given the exceptions to the general tendency that we've discussed in sections 4 and 5, why I haven't accepted that some properly Chisholmed refinement of the sensitivity requirement (which has as instances of it convincing instances of AP's first premise) is necessary for knowledge.
- 12 Though this statement of Nozick's account of knowledge is rough, that will not affect my treatment, which would apply equally well to Nozick's full account. I've skipped entirely Nozick's fourth condition for knowledge, but I believe this fourth

condition to be redundant, anyway: It automatically holds whenever true belief is present. Also, as I've already noted, Nozick takes account of the method of belief formation in his final version of the third condition. The same thing happens with the fourth.

- 13 At pp. 205–6 Nozick admits this appeal, and later he writes, "Thus, if our notion of knowledge was as strong as we naturally tend to think (namely, closed under known logical implication) then the skeptic would be right. (But why do we naturally think this? Further exploration and explanation is needed of the intuitive roots of the natural assumption that knowledge is closed under known logical implication)" (p. 242).

Nozick is quite hard on anti-skeptics who choose rather to deny the first premise; he writes: "The skeptic asserts we do not know his possibilities don't obtain, and he is right. Attempts to avoid skepticism by claiming we do know these things are bound to fail. The skeptic's possibilities make us uneasy because, as we deeply realize, we do not know they don't obtain; it is not surprising that attempts to show we do know these things leave us suspicious, strike us even as bad faith" (p. 201). But similar remarks could be made about Nozick. As Nozick himself admits, the second premise has its own intuitive appeal. So why not say that what we "deeply realize" is that if you don't know that you're not a BIV, then you don't know you have hands, and that the skeptic is right about *this*? Nozick's denial of the second premise leaves me about as "suspicious" as does a denial of the first, and though Nozick's denial doesn't strike me as an instance of bad faith, denials of the first premise seem no better candidates for that charge.

- 14 What are Nozick's grounds for rejecting the second premise? Nozick notes that the premise is an instance of a very general principle to the effect that knowledge is closed under known implication (see note 18, below). After admitting that the closure principle *seems* true (pp. 205–6), Nozick claims that it's wrong, and his reasons for this claim are made entirely from within his analysis of knowledge: Given his analysis, knowledge won't be closed (see especially pp. 206–8). So Nozick is relying on his analysis to show us that the second premise is false despite its intuitive appeal. And indeed, Nozick has developed and defended his analysis of knowledge (in part I of chapter 3) before he applies it to the issue of skepticism (in part 2).
- 15 Unfortunately, what is perhaps the most powerful attack on Nozick's theory of knowledge, made by Saul Kripke in lectures, circa 1985, has not, to the best of my knowledge, found its way into print. For those interested in critical literature on Nozick, a good place to start is with Forbes 1984 and several of the essays in Luper-Foy 1987. For still further read-

ing, Luper-Foy 1987 contains an excellent bibliography.

- 16 As remarked in note 14, Nozick depends heavily on the independent plausibility of this analysis to provide the momentum for his treatment of AI.
- 17 And, of course, such conditionals can be used to make all manner of other comparisons: comparative strength of the epistemic positions of two *different subjects* with respect to the same proposition or with respect to different propositions, the strength of the epistemic position of a subject with respect to one proposition in one situation as compared with that same subject's epistemic position with respect to a different proposition in a different situation, etc.
- 18 As is well known, instances of AI's second premise are often instances of the principle that knowledge is closed under known logical implication: $Kp \ \& \ K(p \text{ entails } q) \rightarrow Kq$. (In the next paragraph I explain why this is not always the case, at least when the closure principle isn't strengthened as there described.) As is also well known, there are exceptions to the principle so formulated, and it might take a lot of tinkering to get it exactly right. But, as Nozick, the arch denier of closure, puts it, "We would be ill-advised, however, to quibble over the details of P [the principle that knowledge is closed under known logical implication]. Although these details are difficult to get straight, it will continue to appear that something like P is correct" (1981, p. 205). Nozick goes on to claim that this appearance is deceiving. I believe that something like P is correct, but that doesn't compete with my present account of AI's second premise: When a conditional is an instance of the properly formulated closure principle, the relevant comparative fact involving strength of epistemic position holds. See Brueckner 1985 for arguments that the denial of knowledge closure principles "is not a fruitful anti-skeptical project" (p. 112).

While restrictions will have to be put on the closure principle that will weaken it in certain respects, there may be other respects in which it can be strengthened. Some instances of AI's second premise are convincing even though H is compatible with O . For instance, the BIV hypothesis seems to undermine my putative knowledge of *I'm in Houston* as well as of *I have hands*, but, of course, that I'm a bodiless BIV is compatible with my being in Houston. Perhaps if S is to know that P , then S must know that not- Q for any Q (but here restrictions must be added) such that if Q were true, S would not know that P . Thus, the range of Q s that must be known not to obtain may be broadened so as to include not only propositions that are incompatible with P , but also others such that if they were the case, then S wouldn't know that P . Those Q s that are incompatible with P itself will then be seen as

special cases of those that are at odds with S 's knowing that P . Barry Stroud discusses a stronger closure principle such as this in his 1984 (pp. 25–30).

- 19 Introducing a skeptical hypothesis into a conversation in any number of ways other than in attributions and denials of knowledge can seem to raise the standards for knowledge. For instance, instead of arguing, "You don't know that the paper isn't mistaken about the result of last night's game; therefore, you don't know that the Bulls won," a skeptic may urge, "Consider this proposition: The newspaper is mistaken about who won the game. Now, keeping that proposition clearly in mind, answer me this: Do you *really* know that the Bulls won?" Of course, not just *any* mention of a skeptical hypothesis seems to trigger the mechanism for raising the standards of knowledge I'm about to articulate.
- 20 Again, I'm here assuming a skeptic-friendly version of contextualism. See the second important point made at the end of section 2.
- 21 But for cases in which it seems one *can* truthfully say "S knows that not- H ," despite the fact that S 's belief that not- H is insensitive, see chapter 3, section J ("Low-Strength Claims to Know that Skeptical Hypotheses Do Not Obtain") of my 1990. In such cases, given certain features of the conversational situation, the Rule of Sensitivity does not operate. These constitute exceptions to the rule that one cannot truthfully call an insensitive belief knowledge. As I explain there, I welcome these exceptions, and would actually be a bit worried if there weren't such exceptions. For it's a feature of my treatment of AI that we do know skeptical hypotheses to be false according to low epistemic standards. I would find it a bit embarrassing if we could never *claim* to have such knowledge by means of simple knowledge attributions, and I'm reassured by the result that in special conversational circumstances, it seems we *can* truthfully claim to know that not- H , despite the fact that our belief that not- H is insensitive.
- 22 This is called the "Moorean" solution because Moore responded in this way to the dream argument. It's far from certain that Moore would have so responded to other instances of AI that utilize different skeptical hypotheses.
- 23 See the first paragraph of note 13, above.
- 24 This is the basic line Unger takes in his defense of bold skepticism in his 1975; see especially pp. 50–4. Stroud, though not himself advocating bold skepticism, does seek to defend the bold skeptic along these lines in ch. 2 of his 1984; see especially pp. 55–82.
- 25 By contrast, our new contextualist solution attributes the apparent truth of (1) to (1)'s *truth* (and not just its warranted assertability) at the very standards its assertion invokes.

- 26 For my own part, for reasons I can't go into here, I think the resulting Moorean position would be slightly more defensible; thus, if I had to reject contextualism and adopt a straightforward solution, I'd be a Moorean.
- 27 Stroud thus claims that on the skeptic's conception of our practices, we operate under certain "practical constraints" (1984, p. 75) in our everyday uses of "know", and asserts that our standards for saying we know vary from case to case (pp. 65–6). Thus, on the skeptic's conception, the standards for ascribing knowledge that we employ in everyday use depend upon our "aims and interests at the moment" (p. 65). According to contextualism, these varying standards reflect a corresponding variation in the truth conditions for attributions of knowledge. But on Stroud's skeptic's conception, when we ascribe knowledge in everyday situations, we are typically saying something literally false, although "the exigencies of action" justify these false attributions. The best exploration of this type of idea is provided by Unger in his 1984.
- 28 Going back to the bold skeptic's first problem, note that all this maneuvering can be mimicked by the Moorean, who can also hold that a Rule of Sensitivity governs the warranted assertability conditions of knowledge ascriptions. Like the bold skeptic, the Moorean can hold that the truth conditions of such attributions of knowledge remain invariant, but in the Moorean's hands, these constant epistemic standards will be meetably low.
- 29 Readers of Unger's 1984 will see the strong influence of that excellent book on my procedure here, though I come to very different conclusions than he does in that work. (But see his more recent 1986.)
- 30 See Stroud (1984, p. 40), who in turn borrowed the example from elsewhere.
- 31 Of course, skeptics are free to refuse this help and propose other solutions. Like practically any claim to have provided the best explanation of something, my claim here is hostage to the possible future development of a better explanation coming along.
- 32 Well, little reason. In his 1984, as part of his case for his relativist conclusion that there's no fact of the matter as to whether contextualism or skeptical invariantism is correct, Unger tries to balance this relative disadvantage of skeptical invariantism against contextualism's relative disadvantage that it does not make the truth conditions of knowledge attributions appropriately independent from the current intents and interests of those who happen to be speaking on a given occasion (p. 37). In part 3 of my 1992, I argue that contextualism can handle the most serious consequences one might suspect would follow from this lack of independence. Whatever independence concerns might remain with contextualism seem quite swamped by the cost of the bold skeptic's solution, which, as I've here argued, is quite high indeed.
- In his review of Unger 1984, Brueckner, relating the advantages of invariantism, writes, "In particular, speakers' intuitions concerning the correct use of 'know' seem to conform to the closure principle for knowledge asserted by the invariantist yet denied by the contextualist" (1986, p. 512). If invariantism, but not contextualism, upheld closure, I would take this to be a very important advantage for invariantism – perhaps even weighty enough to make the contest between the two theories interesting. But, as I've argued, contextualism need not, and, properly developed, does not, take an implausible stand on the issue of closure. (See section 8 and especially note 18, above.)
- 33 I discuss this in section II.B of my 1989.

References

- Brueckner, Anthony I., 1985. "Skepticism and Epistemic Closure," *Philosophical Topics* 13, pp. 89–117.
- , 1986. "Review of Unger, *Philosophical Relativity*," *Journal of Philosophy* 83, pp. 509–17.
- Cohen, Stewart, 1988. "How to Be a Fallibilist," *Philosophical Perspectives* 2, pp. 91–123.
- DeRose, Keith, 1989. "Reid's Anti-Sensationalism and His Realism," *Philosophical Review* 98, pp. 313–48.
- , 1990. "Knowledge, Epistemic Possibility, and Scepticism," Ph.D. diss., University of California, Los Angeles.
- , 1991. "Epistemic Possibilities," *Philosophical Review* 100, pp. 581–605.
- , 1992. "Contextualism and Knowledge Attributions," *Philosophy and Phenomenological Research* 52, pp. 913–29.
- Forbes, Graeme, 1984. "Nozick on Scepticism," *Philosophical Quarterly* 34, pp. 43–52.
- Goldman, Alvin I. 1976. "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73, pp. 771–91.
- Luper-Foy, Stephen (ed.), 1987. *The Possibility of Knowledge: Nozick and his Critics* (Totowa, NJ: Rowman and Littlefield).
- Nozick, Robert, 1981. *Philosophical Explanations* (Cambridge, MA: Harvard University Press).
- Reid, Thomas, 1895. *The Works of Thomas Reid*, 8th edn, ed. William Hamilton (Edinburgh: James Thin).
- Stine, Gail C., 1976. "Skepticism, Relevant Alternatives, and Deductive Closure," *Philosophical Studies* 29, pp. 249–61.
- Stroud, Barry, 1984. *The Significance of Philosophical Scepticism* (Oxford: Oxford University Press).

Keith DeRose

Unger, Peter, 1975. *Ignorance: A Case for Scepticism* (Oxford: Oxford University Press).
——, 1984. *Philosophical Relativity* (Minneapolis: Uni-

versity of Minnesota Press).
——, 1986. "The Cone Model of Knowledge," *Philosophical Topics* 14, pp. 125–78.

Elusive Knowledge

David Lewis

We know a lot. I know what food penguins eat, I know that phones used to ring, but nowadays squeal, when someone calls up. I know that Essendon won the 1993 Grand Final. I know that here is a hand, and here is another.

We have all sorts of everyday knowledge, and we have it in abundance. To doubt that would be absurd. At any rate, to doubt it in any serious and lasting way would be absurd; and even philosophical and temporary doubt, under the influence of argument, is more than a little peculiar. It is a Moorean fact that we know a lot. It is one of those things that we know better than we know the premises of any philosophical argument to the contrary.

Besides knowing a lot that is everyday and trite, I myself think that we know a lot that is interesting and esoteric and controversial. We know a lot about things unseen: tiny particles and pervasive fields, not to mention one another's underwear. Sometimes we even know what an author meant by his writings. But on these questions, let us agree to disagree peacefully with the champions of 'post-knowledgeism'. The most trite and ordinary parts of our knowledge will be problem enough.

For no sooner do we engage in epistemology – the systematic philosophical examination of knowledge – than we meet a compelling argument that we know next to nothing. The sceptical argument is nothing new or fancy. It is just this: it seems as if knowledge must be by definition infallible. If you claim that S knows that *P*, and yet you grant that S

cannot eliminate a certain possibility in which not-*P*, it certainly seems as if you have granted that S does not after all know that *P*. To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just *sounds* contradictory.

Blind Freddy can see where this will lead. Let your paranoid fantasies rip – CIA plots, hallucinogens in the tap water, conspiracies to deceive, old Nick himself – and soon you find that uneliminated possibilities of error are everywhere. Those possibilities of error are far-fetched, of course, but possibilities all the same. They bite into even our most everyday knowledge. We never have infallible knowledge.

Never well, hardly ever. Some say we have infallible knowledge of a few simple, axiomatic necessary truths, and of our own present experience. They say that I simply cannot be wrong that a part of a part of something is itself a part of that thing; or that it seems to me now (as I sit here at the keyboard) exactly as if I am hearing clicking noises on top of a steady whirring. Some say so. Others deny it. No matter; let it be granted, at least for the sake of the argument. It is not nearly enough. If we have only that much infallible knowledge, yet knowledge is by definition infallible, then we have very little knowledge indeed – not the abundant everyday knowledge we thought we had. That is still absurd.

So we know a lot; knowledge must be infallible; yet we have fallible knowledge or none (or next to none). We are caught between the rock of fallibilism and the whirlpool of scepticism. Both are mad!

Yet fallibilism is the less intrusive madness. It demands less frequent corrections of what we want

to say. So, if forced to choose, I choose fallibilism. (And so say all of us.) We can get used to it, and some of us have done. No joy there – we know that people can get used to the most crazy philosophical sayings imaginable. If you are a contented fallibilist, I implore you to be honest, be naive, hear it afresh. 'He knows, yet he has not eliminated all possibilities of error.' Even if you've numbed your ears, doesn't this overt, explicit fallibilism *still* sound wrong?

Better fallibilism than scepticism; but it would be better still to dodge the choice. I think we can. We will be alarmingly close to the rock, and also alarmingly close to the whirlpool, but if we steer with care, we can – just barely – escape them both.

Maybe epistemology is the culprit. Maybe this extraordinary pastime robs us of our knowledge. Maybe we do know a lot in daily life; but maybe when we look hard at our knowledge, it goes away. But only when we look at it harder than the sane ever do in daily life; only when we let our paranoid fantasies rip. That is when we are forced to admit that there always are uneliminated possibilities of error, so that we have fallible knowledge or none.

Much that we say is context-dependent, in simple ways or subtle ways. Sample: 'it's evening' is truly said when, and only when, it is said in the evening. Subtle: it could well be true, and not just by luck, that Essendon played rottenly, the Easybeats played brilliantly, yet Essendon won. Different contexts evoke different standards of evaluation. Talking about the Easybeats we apply lax standards, else we could scarcely distinguish their better days from their worse ones. In talking about Essendon, no such laxity is required. Essendon won because play that is rotten by demanding standards suffices to beat play that is brilliant by lax standards.

Maybe ascriptions of knowledge are subtly context-dependent, and maybe epistemology is a context that makes them go false. Then epistemology would be an investigation that destroys its own subject matter. If so, the sceptical argument might be flawless, when we engage in epistemology – and only then!¹

If you start from the ancient idea that justification is the mark that distinguishes knowledge from mere opinion (even true opinion), then you well might conclude that ascriptions of knowledge are context-dependent because standards for adequate justification are context-dependent. As follows: opinion, even if true, deserves the name of know-

ledge only if it is adequately supported by reasons; to deserve that name in the especially demanding context of epistemology, the arguments from supporting reasons must be especially watertight; but the special standards of justification that this special context demands never can be met (well, hardly ever). In the strict context of epistemology we know nothing, yet in laxer contexts we know a lot.

But I myself cannot subscribe to this account of the context-dependence of knowledge, because I question its starting point. I don't agree that the mark of knowledge is justification.² First, because justification is not sufficient: your true opinion that you will lose the lottery isn't knowledge, whatever the odds. Suppose you know that it is a fair lottery with one winning ticket and many losing tickets, and you know how many losing tickets there are. The greater the number of losing tickets, the better is your justification for believing you will lose. Yet there is no number great enough to transform your fallible opinion into knowledge – after all, you just might win. No justification is good enough – or none short of a watertight deductive argument, and all but the sceptics will agree that this is too much to demand.³

Second, because justification is not always necessary. What (non-circular) argument supports our reliance on perception, on memory, and on testimony?⁴ And yet we do gain knowledge by these means. And sometimes, far from having supporting arguments, we don't even know how we know. We once had evidence, drew conclusions, and thereby gained knowledge; now we have forgotten our reasons, yet still we retain our knowledge. Or we know the name that goes with the face, or the sex of the chicken, by relying on subtle visual cues, without knowing what those cues may be.

The link between knowledge and justification must be broken. But if we break that link, then it is not – or not entirely, or not exactly – by raising the standards of justification that epistemology destroys knowledge. I need some different story.

To that end, I propose to take the infallibility of knowledge as my starting point.⁵ Must infallibilist epistemology end in scepticism? Not quite. Wait and see. Anyway, here is the definition. Subject *S* knows proposition *P* iff *P* holds in every possibility left uneliminated by *S*'s evidence; equivalently, iff *S*'s evidence eliminates every possibility in which not-*P*.

The definition is short, the commentary upon it is longer. In the first place, there is the proposition, *P*. What I choose to call 'propositions' are individuated coarsely, by necessary equivalence. For instance, there is only one necessary proposition. It holds in every possibility; hence in every possibility left uneliminated by *S*'s evidence, no matter who *S* may be and no matter what his evidence may be. So the necessary proposition is known always and everywhere. Yet this known proposition may go unrecognised when presented in impenetrable linguistic disguise, say as the proposition that every even number is the sum of two primes. Likewise, the known proposition that I have two hands may go unrecognised when presented as the proposition that the number of my hands is the least number *n* such that every even number is the sum of *n* primes. (Or if you doubt the necessary existence of numbers, switch to an example involving equivalence by logic alone.) These problems of disguise shall not concern us here. Our topic is modal, not hyperintensional, epistemology.⁶

Next, there are the possibilities. We needn't enter here into the question whether these are concreta, abstract constructions, or abstract simples. Further, we needn't decide whether they must always be maximally specific possibilities, or whether they need only be specific enough for the purpose at hand. A possibility will be specific enough if it cannot be split into sub-cases in such a way that anything we have said about possibilities, or anything we are going to say before we are done, applies to some sub-cases and not to others. For instance, it should never happen that proposition *P* holds in some but not all sub-cases; or that some but not all sub-cases are eliminated by *S*'s evidence.

But we do need to stipulate that they are not just possibilities as to how the whole world is; they also include possibilities as to which part of the world is oneself, and as to when it now is. We need these possibilities *de se et nunc* because the propositions that may be known include propositions *de se et nunc*.⁷ Not only do I know that there are hands in this world somewhere and somewhen. I know that I have hands, or anyway I have them *now*. Such propositions aren't just made true or made false by the whole world once and for all. They are true for some of us and not for others, or true at some times and not others, or both.

Further, we cannot limit ourselves to 'real' possibilities that conform to the actual laws of nature,

and maybe also to actual past history. For propositions about laws and history are contingent, and may or may not be known.

Neither can we limit ourselves to 'epistemic' possibilities for *S* – possibilities that *S* does not know not to obtain. That would drain our definition of content. Assume only that knowledge is closed under strict implication. (We shall consider the merits of this assumption later.) Remember that we are not distinguishing between equivalent propositions. Then knowledge of a conjunction is equivalent to knowledge of every conjunct. *P* is the conjunction of all propositions not-*W*, where *W* is a possibility in which not-*P*. That suffices to yield an equivalence: *S* knows that *P* iff, for every possibility *W* in which not-*P*, *S* knows that not-*W*. Contraposing and cancelling a double negation: iff every possibility which *S* does not know not to obtain is one in which *P*. For short: iff *P* holds throughout *S*'s epistemic possibilities. Yet to get this far, we need no substantive definition of knowledge at all! To turn this into a substantive definition, in fact the very definition we gave before, we need to say one more thing: *S*'s epistemic possibilities are just those possibilities that are uneliminated by *S*'s evidence.

So, next, we need to say what it means for a possibility to be eliminated or not. Here I say that the uneliminated possibilities are those in which the subject's entire perceptual experience and memory are just as they actually are. There is one possibility that actually obtains (for the subject and at the time in question); call it *actuality*. Then a possibility *W* is *uneliminated* iff the subject's perceptual experience and memory in *W* exactly match his perceptual experience and memory in actuality. (If you want to include other alleged forms of basic evidence, such as the evidence of our extrasensory faculties, or an innate disposition to believe in God, be my guest. If they exist, they should be included. If not, no harm done if we have included them conditionally.)

Note well that we do not need the 'pure sense-datum language' and the 'incorrigible protocol statements' that for so long bedevilled foundationalist epistemology. It matters not at all whether there are words to capture the subject's perceptual and memory evidence, nothing more and nothing less. If there are such words, it matters not at all whether the subject can hit upon them. The given does not consist of basic axioms to serve as premises in subsequent arguments. Rather, it consists of a match between possibilities.

When perceptual experience E (or memory) eliminates a possibility W , that is not because the propositional content of the experience conflicts with W . (Not even if it is the narrow content.) The propositional content of our experience could, after all, be false. Rather, it is the existence of the experience that conflicts with W . W is a possibility in which the subject is not having experience E . Else we would need to tell some fishy story of how the experience has some sort of infallible, ineffable, purely phenomenal propositional content... Who needs that? Let E have propositional content P . Suppose even – something I take to be an open question – that E is, in some sense, fully characterized by P . Then I say that E eliminates W iff W is a possibility in which the subject's experience or memory has content different from P . I do *not* say that E eliminates W iff W is a possibility in which P is false.

Maybe not every kind of sense perception yields experience; maybe, for instance, the kinaesthetic sense yields not its own distinctive sort of sense-experience but only spontaneous judgements about the position of one's limbs. If this is true, then the thing to say is that kinaesthetic evidence eliminates all possibilities except those that exactly resemble actuality with respect to the subject's spontaneous kinaesthetic judgements. In saying this, we would treat kinaesthetic evidence more on the model of memory than on the model of more typical senses.

Finally, we must attend to the word 'every'. What does it mean to say that every possibility in which not- P is eliminated? An idiom of quantification, like 'every', is normally restricted to some limited domain. If I say that every glass is empty, so it's time for another round, doubtless I and my audience are ignoring most of all the glasses there are in the whole wide world throughout all of time. They are outside the domain. They are irrelevant to the truth of what was said.

Likewise, if I say that every uneliminated possibility is one in which P , or words to that effect, I am doubtless ignoring some of all the uneliminated alternative possibilities that there are. They are outside the domain, they are irrelevant to the truth of what was said.

But, of course, I am not entitled to ignore just any possibility I please. Else true ascriptions of knowledge, whether to myself or to others, would be cheap indeed. I may properly ignore some uneliminated possibilities; I may not properly ignore

others. Our definition of knowledge requires a *sotto voce* proviso. S knows that P iff S 's evidence eliminates every possibility in which not- P – Pss! – except for those possibilities that we are properly ignoring.

Unger suggests an instructive parallel.⁸ Just as P is known iff there are no uneliminated possibilities of error, so likewise a surface is flat iff there are no bumps on it. We must add the proviso: Pss! – except for those bumps that we are properly ignoring. Else we will conclude, absurdly, that nothing is flat. (Simplify by ignoring departures from flatness that consist of gentle curvature.)

We can restate the definition. Say that we *presuppose* proposition Q iff we ignore all possibilities in which not- Q . To close the circle: we *ignore* just those possibilities that falsify our presuppositions. *Proper* presupposition corresponds, of course, to proper ignoring. Then S knows that P iff S 's evidence eliminates every possibility in which not- P – Pss! – except for those possibilities that conflict with our proper presuppositions.⁹

The rest of (modal) epistemology examines the *sotto voce* proviso. It asks: what may we properly presuppose in our ascriptions of knowledge? Which of all the uneliminated alternative possibilities may not properly be ignored? Which ones are the 'relevant alternatives'? – relevant, that is, to what the subject does and doesn't know?¹⁰ In reply, we can list several rules.¹¹ We begin with three prohibitions: rules to tell us what possibilities we may not properly ignore.

First, there is the *Rule of Actuality*. The possibility that actually obtains is never properly ignored; actuality is always a relevant alternative; nothing false may properly be presupposed. It follows that only what is true is known, wherefore we did not have to include truth in our definition of knowledge. The rule is 'externalist' – the subject himself may not be able to tell what is properly ignored. In judging which of his ignorings are proper, hence what he knows, we judge his success in knowing – not how well he tried.

When the Rule of Actuality tells us that actuality may never be properly ignored, we can ask: *whose* actuality? Ours, when we ascribe knowledge or ignorance to others? Or the subject's? In simple cases, the question is silly. (In fact, it sounds like the sort of pernicious nonsense we would expect from someone who mixes up what is true with what is believed.) There is just one actual world, we the ascribers live in that world, the subject lives

there too, so the subject's actuality is the same as ours.

But there are other cases, less simple, in which the question makes perfect sense and needs an answer. Someone may or may not know who he is; someone may or may not know what time it is. Therefore I insisted that the propositions that may be known must include propositions *de se et nunc*; and likewise that the possibilities that may be eliminated or ignored must include possibilities *de se et nunc*. Now we have a good sense in which the subject's actuality may be different from ours. I ask today what Fred knew yesterday. In particular, did he then know who he was? Did he know what day it was? Fred's actuality is the possibility *de se et nunc* of being Fred on September 19th at such-and-such possible world; whereas my actuality is the possibility *de se et nunc* of being David on September 20th at such-and-such world. So far as the world goes, there is no difference: Fred and I are worldmates, his actual world is the same as mine. But when we build subject and time into the possibilities *de se et nunc*, then his actuality yesterday does indeed differ from mine today.

What is more, we sometimes have occasion to ascribe knowledge to those who are off at other possible worlds. I didn't read the newspaper yesterday. What would I have known if I had read it? More than I do in fact know. (More and less: I do in fact know that I left the newspaper unread, but if I had read it, I would not have known that I had left it unread.) I—who-did-not-read-the-newspaper am here at this world, ascribing knowledge and ignorance. The subject to whom I am ascribing that knowledge and ignorance, namely I-as-I-would-have-been-if-I-had-read-the-newspaper, is at a different world. The worlds differ in respect at least of a reading of the newspaper. Thus the ascriber's actual world is not the same as the subject's. (I myself think that the ascriber and the subject are two different people: the subject is the ascriber's otherworldly counterpart. But even if you think the subject and the ascriber are the same identical person, you must still grant that this person's actuality *qua* subject differs from his actuality *qua* ascriber.)

Or suppose we ask modal questions about the subject: what must he have known, what might he have known? Again we are considering the subject as he is not here, but off at other possible worlds. Likewise if we ask questions about knowledge of knowledge: what does he (or what do we) know that he knows?

So the question 'whose actuality?' is not a silly question after all. And when the question matters, as it does in the cases just considered, the right answer is that it is the subject's actuality, not the ascriber's, that never can be properly ignored.

Next, there is the *Rule of Belief*. A possibility that the subject believes to obtain is not properly ignored, whether or not he is right to so believe. Neither is one that he ought to believe to obtain — one that evidence and arguments justify him in believing — whether or not he does so believe.

That is rough. Since belief admits of degree, and since some possibilities are more specific than others, we ought to reformulate the rule in terms of degree of belief, compared to a standard set by the unspecificity of the possibility in question. A possibility may not be properly ignored if the subject gives it, or ought to give it, a degree of belief that is sufficiently high, and high not just because the possibility in question is unspecific.

How high is 'sufficiently high'? That may depend on how much is at stake. When error would be especially disastrous, few possibilities may be properly ignored. Then even quite a low degree of belief may be 'sufficiently high' to bring the Rule of Belief into play. The jurors know that the accused is guilty only if his guilt has been proved beyond reasonable doubt.¹²

Yet even when the stakes are high, some possibilities still may be properly ignored. Disastrous though it would be to convict an innocent man, still the jurors may properly ignore the possibility that it was the dog, marvellously well-trained, that fired the fatal shot. And, unless they are ignoring other alternatives more relevant than that, they may rightly be said to know that the accused is guilty as charged. Yet if there had been reason to give the dog hypothesis a slightly less negligible degree of belief — if the world's greatest dog-trainer had been the victim's mortal enemy — then the alternative would be relevant after all.

This is the only place where belief and justification enter my story. As already noted, I allow justified true belief without knowledge, as in the case of your belief that you will lose the lottery. I allow knowledge without justification, in the cases of face recognition and chicken sexing. I even allow knowledge without belief, as in the case of the timid student who knows the answer but has no confidence that he has it right, and so does not believe what he knows.¹³ Therefore any proposed

converse to the Rule of Belief should be rejected. A possibility that the subject does not believe to a sufficient degree, and ought not to believe to a sufficient degree, may nevertheless be a relevant alternative and not properly ignored.

Next, there is the *Rule of Resemblance*. Suppose one possibility saliently resembles another. Then if one of them may not be properly ignored, neither may the other. (Or rather, we should say that if one of them may not properly be ignored *in virtue of rules other than this rule*, then neither may the other. Else nothing could be properly ignored; because enough little steps of resemblance can take us from anywhere to anywhere.) Or suppose one possibility saliently resembles two or more others, one in one respect and another in another, and suppose that each of these may not properly be ignored (in virtue of rules other than this rule). Then these resemblances may have an additive effect, doing more together than any one of them would separately.

We must apply the Rule of Resemblance with care. Actuality is a possibility uneliminated by the subject's evidence. Any other possibility *W* that is likewise uneliminated by the subject's evidence thereby resembles actuality in one salient respect; namely, in respect of the subject's evidence. That will be so even if *W* is in other respects very dissimilar to actuality — even if, for instance, it is a possibility in which the subject is radically deceived by a demon. Plainly, we dare not apply the Rules of Actuality and Resemblance to conclude that any such *W* is a relevant alternative — that would be capitulation to scepticism. The Rule of Resemblance was never meant to apply to *this* resemblance! We seem to have an *ad hoc* exception to the Rule, though one that makes good sense in view of the function of attributions of knowledge. What would be better, though, would be to find a way to reformulate the Rule so as to get the needed exception without *ad hoc*ery. I do not know how to do this.

It is the Rule of Resemblance that explains why you do not know that you will lose the lottery, no matter what the odds are against you and no matter how sure you should therefore be that you will lose. For every ticket, there is the possibility that it will win. These possibilities are saliently similar to one another: so either every one of them may be properly ignored, or else none may. But one of them may not properly be ignored: the one that actually obtains.

The Rule of Resemblance also is the rule that solves the Gettier problems: other cases of justified true belief that are not knowledge.¹⁴

(1) I think that Nogot owns a Ford, because I have seen him driving one; but unbeknownst to me he does not own the Ford he drives, or any other Ford. Unbeknownst to me, Havit does own a Ford, though I have no reason to think so because he never drives it, and in fact I have often seen him taking the tram. My justified true belief is that one of the two owns a Ford. But I do not know it; I am right by accident. Diagnosis: I do not know, because I have not eliminated the possibility that Nogot drives a Ford he does not own whereas Havit neither drives nor owns a car. This possibility may not properly be ignored. Because, first, actuality may not properly be ignored; and, second, this possibility saliently resembles actuality. It resembles actuality perfectly so far as Nogot is concerned; and it resembles actuality well so far as Havit is concerned, since it matches actuality both with respect to Havit's careless habits and with respect to the general correlation between careless habits and carelessness. In addition, this possibility saliently resembles a third possibility: one in which Nogot drives a Ford he owns while Havit neither drives nor owns a car. This third possibility may not properly be ignored, because of the degree to which it is believed. This time, the resemblance is perfect so far as Havit is concerned, rather good so far as Nogot is concerned.

(2) The stopped clock is right twice a day. It says 4:39, as it has done for weeks. I look at it at 4:39; by luck I pick up a true belief. I have ignored the uneliminated possibility that I looked at it at 4:22 while it was stopped saying 4:39. That possibility was not properly ignored. It resembles actuality perfectly so far as the stopped clock goes.

(3) Unbeknownst to me, I am travelling in the land of the bogus barns; but my eye falls on one of the few real ones. I don't know that I am seeing a barn, because I may not properly ignore the possibility that I am seeing yet another of the abundant bogus barns. This possibility saliently resembles actuality in respect of the abundance of bogus barns, and the scarcity of real ones, hereabouts.

(4) Donald is in San Francisco, just as I have every reason to think he is. But, bent on deception, he is writing me letters and having them posted to me by his accomplice in Italy. If I had seen the phoney letters, with their Italian stamps and post-

marks, I would have concluded that Donald was in Italy. Luckily, I have not yet seen any of them. I ignore the uneliminated possibility that Donald has gone to Italy and is sending me letters from there. But this possibility is not properly ignored, because it resembles actuality both with respect to the fact that the letters are coming to me from Italy and with respect to the fact that those letters come, ultimately, from Donald. So I don't know that Donald is in San Francisco

Next, there is the *Rule of Reliability*. This time, we have a presumptive rule about what *may* be properly ignored; and it is by means of this rule that we capture what is right about causal or reliabilist theories of knowing. Consider processes whereby information is transmitted to us: perception, memory, and testimony. These processes are fairly reliable.¹⁵ Within limits, we are entitled to take them for granted. We may properly presuppose that they work without a glitch in the case under consideration. Defeasibly – very defeasibly! – a possibility in which they fail may properly be ignored.

My visual experience, for instance, depends causally on the scene before my eyes, and what I believe about the scene before my eyes depends in turn on my visual experience. Each dependence covers a wide and varied range of alternatives.¹⁶ Of course, it is possible to hallucinate – even to hallucinate in such a way that all my perceptual experience and memory would be just as they actually are. That possibility never can be eliminated. But it can be ignored. And if it is properly ignored – as it mostly is – then vision gives me knowledge. Sometimes, though, the possibility of hallucination is not properly ignored; for sometimes we really do hallucinate. The Rule of Reliability may be defeated by the Rule of Actuality. Or it may be defeated by the Rules of Actuality and of Resemblance working together, in a Gettier problem: if I am not hallucinating, but unbeknownst to me I live in a world where people mostly do hallucinate and I myself have only narrowly escaped, then the uneliminated possibility of hallucination is too close to actuality to be properly ignored.

We do not, of course, presuppose that nowhere ever is there a failure of, say, vision. The general presupposition that vision is reliable consists, rather, of a standing disposition to presuppose, concerning whatever particular case may be under consideration, that we have no failure in that case.

In similar fashion, we have two permissive *Rules of Method*. We are entitled to presuppose again, very defeasibly – that a sample is representative; and that the best explanation of our evidence is the true explanation. That is, we are entitled properly to ignore possible failures in these two standard methods of non-deductive inference. Again, the general rule consists of a standing disposition to presuppose reliability in whatever particular case may come before us.

Yet another permissive rule is the *Rule of Conservatism*. Suppose that those around us normally do ignore certain possibilities, and it is common knowledge that they do (They do, they expect each other to, they expect each other to expect each other to, ...) Then again, very defeasibly! – these generally ignored possibilities may properly be ignored. We are permitted, defeasibly, to adapt the usual and mutually expected presuppositions of those around us.

(It is unclear whether we need all four of these permissive rules. Some might be subsumed under others. Perhaps our habits of treating samples as representative, and of inferring to the best explanation, might count as normally reliable processes of transmission of information. Or perhaps we might subsume the Rule of Reliability under the Rule of Conservatism, on the ground that the reliable processes whereby we gain knowledge are familiar, are generally relied upon, and so are generally presupposed to be normally reliable. Then the only extra work done by the Rule of Reliability would be to cover less familiar – and merely hypothetical? – reliable processes, such as processes that relied on extrasensory faculties. Likewise, *mutatis mutandis*, we might subsume the Rules of Method under the Rule of Conservatism. Or we might instead think to subsume the Rule of Conservatism under the Rule of Reliability, on the ground that what is generally presupposed tends for the most part to be true, and the reliable processes whereby this is so are covered already by the Rule of Reliability. Better redundancy than incompleteness, though. So, leaving the question of redundancy open, I list all four rules.)

Our final rule is the *Rule of Attention*. But it is more a triviality than a rule. When we say that a possibility is properly ignored, we mean exactly that; we do not mean that it *could have been* properly ignored. Accordingly, a possibility not ignored

at all is *ipso facto* not properly ignored. What is and what is not being ignored is a feature of the particular conversational context. No matter how far-fetched a certain possibility may be, no matter how properly we might have ignored it in some other context, if in *this* context we are not in fact ignoring it but attending to it, then for us now it is a relevant alternative. It is in the contextually determined domain. If it is an uneliminated possibility in which not-*P*, then it will do as a counter-example to the claim that *P* holds in every possibility left uneliminated by *S*'s evidence. That is, it will do as a counter-example to the claim that *S* knows that *P*.

Do some epistemology. Let your fantasies rip. Find uneliminated possibilities of error everywhere. Now that you are attending to them, just as I told you to, you are no longer ignoring them, properly or otherwise. So you have landed in a context with an enormously rich domain of potential counter-examples to ascriptions of knowledge. In such an extraordinary context, with such a rich domain, it never can happen (well, hardly ever) that an ascription of knowledge is true. Not an ascription of knowledge to yourself (either to your present self or to your earlier self, untainted by epistemology); and not an ascription of knowledge to others. That is how epistemology destroys knowledge. But it does so only temporarily. The pastime of epistemology does not plunge us forevermore into its special context. We can still do a lot of proper ignoring, a lot of knowing, and a lot of true ascribing of knowledge to ourselves and others, the rest of the time.

What is epistemology all about? The epistemology we've just been doing, at any rate, soon became an investigation of the ignoring of possibilities. But to investigate the ignoring of them was *ipso facto* not to ignore them. Unless this investigation of ours was an altogether atypical sample of epistemology, it will be inevitable that epistemology must destroy knowledge. That is how knowledge is elusive. Examine it, and straightway it vanishes.

Is resistance useless? If you bring some hitherto ignored possibility to our attention, then straightway we are not ignoring it at all, so *a fortiori* we are not properly ignoring it. How can this alteration of our conversational state be undone? If you are persistent, perhaps it cannot be undone – at least not so long as you are around. Even if we go off and play backgammon, and afterward start our

conversation afresh, you might turn up and call our attention to it all over again.

But maybe you called attention to the hitherto ignored possibility by mistake. You only suggested that we ought to suspect the butler because you mistakenly thought him to have a criminal record. Now that you know he does not – that was the *previous* butler – you wish you had not mentioned him at all. You know as well as we do that continued attention to the possibility you brought up impedes our shared conversational purposes. Indeed, it may be common knowledge between you and us that we would all prefer it if this possibility could be dismissed from our attention. In that case we might quickly strike a tacit agreement to speak just as if we were ignoring it; and after just a little of that, doubtless it really would be ignored.

Sometimes our conversational purposes are not altogether shared, and it is a matter of conflict whether attention to some far-fetched possibility would advance them or impede them. What if some far-fetched possibility is called to our attention not by a sceptical philosopher, but by counsel for the defence? We of the jury may wish to ignore it, and wish it had not been mentioned. If we ignored it now, we would bend the rules of cooperative conversation; but we may have good reason to do exactly that. (After all, what matters most to us as jurors is not whether we can truly be said to know; what really matters is what we should believe to what degree, and whether or not we should vote to convict.) We would ignore the far-fetched possibility if we could – but can we? Perhaps at first our attempted ignoring would be make-believe ignoring, or self-deceptive ignoring; later, perhaps, it might ripen into genuine ignoring. But in the meantime, do we know? There may be no definite answer. We are bending the rules, and our practices of context-dependent attributions of knowledge were made for contexts with the rules unent.

If you are still a contented fallibilist, despite my plea to hear the sceptical argument afresh, you will probably be discontented with the Rule of Attention. You will begrudge the sceptic even his very temporary victory. You will claim the right to resist his argument not only in everyday contexts, but even in those peculiar contexts in which he (or some other epistemologist) busily calls your attention to far-fetched possibilities of error. Further, you will claim the right to resist without having to bend any rules of cooperative conversation. I said

that the Rule of Attention was a triviality: that which is not ignored at all is not properly ignored. But the Rule was trivial only because of how I had already chosen to state the *sotto voce* proviso. So you, the contented fallibilist, will think it ought to have been stated differently. Thus, perhaps: 'Psst! – except for those possibilities we *could* properly have ignored.' And then you will insist that those far-fetched possibilities of error that we attend to at the behest of the sceptic are nevertheless possibilities we could properly have ignored. You will say that no amount of attention can, by itself, turn them into relevant alternatives.

If you say this, we have reached a standoff. I started with a puzzle: how can it be, when his conclusion is so silly, that the sceptic's argument is so irresistible? My Rule of Attention, and the version of the proviso that made that Rule trivial, were built to explain how the sceptic manages to sway us – why his argument seems irresistible, however temporarily. If you continue to find it eminently resistible in all contexts, you have no need of any such explanation. We just disagree about the explanandum phenomenon.

I say S knows that *P* iff *P* holds in every possibility left uneliminated by S's evidence – Psst! – except for those possibilities that *we* are properly ignoring. 'We' means: the speaker and hearers of a given context; that is, those of us who are discussing S's knowledge together. It is our ignorings, not S's own ignorings, that matter to what we can truly say about S's knowledge. When we are talking about our own knowledge or ignorance, as epistemologists so often do, this is a distinction without a difference. But what if we are talking about someone else?

Suppose we are detectives; the crucial question for our solution of the crime is whether S already *knew*, when he bought the gun, that he was vulnerable to blackmail. We conclude that he did. *We* ignore various far-fetched possibilities, as hard-headed detectives should. But S does not ignore them. S is by profession a sceptical epistemologist. He never ignores much of anything. If it is our own ignorings that matter to the truth of our conclusion, we may well be right that S already knew. But if it is S's ignorings that matter, then we are wrong, because S never knew much of anything. I say we may well be right; so it is our own ignorings that matter, not S's.

But suppose instead that we are epistemologists considering what S knows. If we are well-informed

about S (or if we are considering a well-enough specified hypothetical case), then if S attends to a certain possibility, we attend to S's attending to it. But to attend to S's attending to it is *ipso facto* to attend to it ourselves. In that case, unlike the case of the detectives, the possibilities we are properly ignoring must be among the possibilities that S himself ignores. We may ignore fewer possibilities than S does, but not more.

Even if S himself is neither sceptical nor an epistemologist, he may yet be clever at thinking up far-fetched possibilities that are uneliminated by his evidence. Then again, we well-informed epistemologists who ask what S knows will have to attend to the possibilities that S thinks up. Even if S's idle cleverness does not lead S himself to draw sceptical conclusions, it nevertheless limits the knowledge that we can truly ascribe to him when attentive to his state of mind. More simply: his cleverness limits his knowledge. He would have known more, had he been less imaginative.¹⁷

Do I claim you can know *P* just by presupposing it?! Do I claim you can know that a possibility *W* does not obtain just by ignoring it? Is that not what my analysis implies, provided that the presupposing and the ignoring are proper? Well, yes. And yet I do not claim it. Or rather, I do not claim it for any specified *P* or *W*. I have to grant, in general, that knowledge just by presupposing and ignoring *is* knowledge; but it is an *especially* elusive sort of knowledge, and consequently it is an unclaimable sort of knowledge. You do not even have to practise epistemology to make it vanish. Simply *mentioning* any particular case of this knowledge, aloud or even in silent thought, is a way to attend to the hitherto ignored possibility, and thereby render it no longer ignored, and thereby create a context in which it is no longer true to ascribe the knowledge in question to yourself or others. So, just as we should think, presuppositions alone are not a basis on which to *claim* knowledge.

In general, when S knows that *P* some of the possibilities in which not-*P* are eliminated by S's evidence and others of them are properly ignored. There are some that can be eliminated, but cannot properly be ignored. For instance, when I look around the study without seeing Possum the cat, I thereby eliminate various possibilities in which Possum is in the study; but had those possibilities not been eliminated, they could not properly have been ignored. And there are other possibilities that

never can be eliminated, but can properly be ignored. For instance, the possibility that Possum is on the desk but has been made invisible by a deceiving demon falls normally into this class (though not when I attend to it in the special context of epistemology).

There is a third class: not-*P* possibilities that might either be eliminated or ignored. Take the far-fetched possibility that Possum has somehow managed to get into a closed drawer of the desk – maybe he jumped in when it was open, then I closed it without noticing him. That possibility could be eliminated by opening the drawer and making a thorough examination. But if uneliminated, it may nevertheless be ignored, and in many contexts that ignoring would be proper. If I look all around the study, but without checking the closed drawers of the desk, I may truly be said to know that Possum is not in the study – or at any rate, there are many contexts in which that may truly be said. But if I did check all the closed drawers, then I would know *better* that Possum is not in the study. My knowledge would be better in the second case because it would rest more on the elimination of not-*P* possibilities, less on the ignoring of them.^{18 19}

Better knowledge is more stable knowledge: it stands more chance of surviving a shift of attention in which we begin to attend to some of the possibilities formerly ignored. If, in our new shifted context, we ask what knowledge we may truly ascribe to our earlier selves, we may find that only the better knowledge of our earlier selves still deserves the name. And yet, if our former ignorings were proper at the time, even the worse knowledge of our earlier selves could truly have been called knowledge in the former context.

Never – well, hardly ever – does our knowledge rest entirely on elimination and not at all on ignoring. So hardly ever is it quite as good as we might wish. To that extent, the lesson of scepticism is right – and right permanently, not just in the temporary and special context of epistemology.²⁰

What is it all for? Why have a notion of knowledge that works in the way I described? (Not a compulsory question. Enough to observe that we do have it.) But I venture the guess that it is one of the messy short-cuts – like satisficing, like having indeterminate degrees of belief – that we resort to because we are not smart enough to live up to really high, perfectly Bayesian, standards of ration-

ality. You cannot maintain a record of exactly which possibilities you have eliminated so far, much as you might like to. It is easier to keep track of which possibilities you have eliminated if you – Psst! – ignore many of all the possibilities there are. And besides, it is easier to list some of the propositions that are true in *all* the uneliminated, unignored possibilities than it is to find propositions that are true in *all and only* the uneliminated, unignored possibilities.

If you doubt that the word ‘know’ bears any real load in science or in metaphysics, I partly agree. The serious business of science has to do not with knowledge *per se*; but rather, with the elimination of possibilities through the evidence of perception, memory, etc., and with the changes that one’s belief system would (or might or should) undergo under the impact of such eliminations. Ascriptions of knowledge to yourself or others are a very sloppy way of conveying very incomplete information about the elimination of possibilities. It is as if you had said:

The possibilities eliminated, whatever else they may also include, at least include all the not-*P* possibilities; or anyway, all of those except for some we are presumably prepared to ignore just at the moment.

The only excuse for giving information about what really matters in such a sloppy way is that at least it is easy and quick! But it *is* easy and quick; whereas giving full and precise information about which possibilities have been eliminated seems to be extremely difficult, as witness the futile search for a ‘pure observation language’. If I am right about how ascriptions of knowledge work, they are a handy but humble approximation. They may yet be indispensable in practice, in the same way that other handy and humble approximations are.

If we analyse knowledge as a modality, as we have done, we cannot escape the conclusion that knowledge is closed under (strict) implication.²¹ Dretske has denied that knowledge is closed under implication; further, he has diagnosed closure as the fallacy that drives arguments for scepticism. As follows: the proposition that I have hands implies that I am not a handless being, and *a fortiori* that I am not a handless being deceived by a demon into thinking that I have hands. So, by the closure principle, the proposition that I know I have hands implies that I know that I know I have hands implies that I know that I am not handless and deceived. But

I don't know that I am not handless and deceived – for how can I eliminate that possibility? So, by *modus tollens*, I don't know that I have hands. Dretske's advice is to resist scepticism by denying closure. He says that although having hands *does* imply not being handless and deceived, yet knowing that I have hands *does not* imply knowing that I am not handless and deceived. I do know the former, I do not know the latter.²²

What Dretske says is close to right, but not quite. Knowledge *is* closed under implication. Knowing that I have hands *does* imply knowing that I am not handless and deceived. Implication preserves truth – that is, it preserves truth in any given, fixed context. But if we switch contexts midway, all bets are off. I say (1) pigs fly; (2) what I just said had fewer than three syllables (true); (3) what I just said had fewer than four syllables (false). So 'less than three' does not imply 'less than four'? No! The context switched midway, the semantic value of the context-dependent phrase 'what I just said' switched with it. Likewise in the sceptical argument the context switched midway, and the semantic value of the context-dependent word 'know' switched with it. The premise 'I know that I have hands' was true in its everyday context, where the possibility of deceiving demons was properly ignored. The mention of that very possibility switched the context midway. The conclusion 'I know that I am not handless and deceived' was false in *its* context, because that was a context in which the possibility of deceiving demons was being mentioned, hence was not being ignored, hence was not being properly ignored. Dretske gets the phenomenon right, and I think he gets the diagnosis of scepticism right; it is just that he misclassifies what he sees. He thinks it is a phenomenon of logic, when really it is a phenomenon of pragmatics. Closure, rightly understood, survives the test. If we evaluate the conclusion for truth not with respect to the context in which it was uttered, but instead with respect to the different context in which the premise was uttered, then truth is preserved. And if, *per impossibile*, the conclusion could have been said in the same unchanged context as the premise, truth would have been preserved.

A problem due to Saul Kripke turns upon the closure of knowledge under implication. *P* implies that any evidence against *P* is misleading. So, by closure, whenever you know that *P*, you know that any evidence against *P* is misleading. And if you know that evidence is misleading, you should pay

it no heed. Whenever we know – and we know a lot, remember – we should not heed any evidence tending to suggest that we are wrong. But that is absurd. Shall we dodge the conclusion by denying closure? I think not. Again, I diagnose a change of context. At first, it was stipulated that *S* knew, whence it followed that *S* was properly ignoring all possibilities of error. But as the story continues, it turns out that there is evidence on offer that points to some particular possibility of error. Then, by the Rule of Attention, that possibility is no longer properly ignored, either by *S* himself or by we who are telling the story of *S*. The advent of that evidence destroys *S*'s knowledge, and thereby destroys *S*'s licence to ignore the evidence lest he be misled.

There is another reason, different from Dretske's, why we might doubt closure. Suppose two or more premises jointly imply a conclusion. Might not someone who is compartmentalized in his thinking – as we all are? – know each of the premises but fail to bring them together in a single compartment? Then might he not fail to know the conclusion? Yes; and I would not like to plead idealization-of-rationality as an excuse for ignoring such cases. But I suggest that we might take not the whole compartmentalized thinker, but rather each of his several overlapping compartments, as our 'subjects'. That would be the obvious remedy if his compartmentalization amounted to a case of multiple personality disorder; but maybe it is right for milder cases as well.²³

A compartmentalized thinker who indulges in epistemology can destroy his knowledge, yet retain it as well. Imagine two epistemologists on a bushwalk. As they walk, they talk. They mention all manner of far-fetched possibilities of error. By attending to these normally ignored possibilities they destroy the knowledge they normally possess. Yet all the while they know where they are and where they are going! How so? The compartment in charge of philosophical talk attends to far-fetched possibilities of error. The compartment in charge of navigation does not. One compartment loses its knowledge, the other retains its knowledge. And what does the entire compartmentalized thinker know? Not an altogether felicitous question. But if we need an answer, I suppose the best thing to say is that *S* knows that *P* iff any one of *S*'s compartments knows that *P*. Then we can say what we would offhand want to say: yes, our philosophical bushwalkers still know their whereabouts.

Context-dependence is not limited to the ignoring and non-ignoring of far-fetched possibilities. Here is another case. Pity poor Bill! He squanders all his spare cash on the pokies, the races, and the lottery. He will be a wage slave all his days. We know he will never be rich. But if he wins the lottery (if he wins big), then he will be rich. Contrapositively: his never being rich, plus other things we know, imply that he will lose. So, by closure, if we know that he will never be rich, we know that he will lose. But when we discussed the case before, we concluded that we cannot know that he will lose. All the possibilities in which Bill loses and someone else wins saliently resemble the possibility in which Bill wins and the others lose; one of those possibilities is actual; so by the Rules of Actuality and of Resemblance, we may not properly ignore the possibility that Bill wins. But there is a loophole: the resemblance was required to be salient. Salience, as well as ignoring, may vary between contexts. Before, when I was explaining how the Rule of Resemblance applied to lotteries, I saw to it that the resemblance between the many possibilities associated with the many tickets was sufficiently salient. But this time, when we were busy pitying poor Bill for his habits and not for his luck, the resemblance of the many possibilities was not so salient. At that point, the possibility of Bill's winning was properly ignored; so then it was true to say that we knew he would never be rich. Afterward I switched the context. I mentioned the possibility that Bill might win, wherefore that possibility was no longer properly ignored. (Maybe there were two separate reasons why it was no longer properly ignored, because maybe I also made the resemblance between the many possibilities more salient.) It was true at first that we knew that Bill would never be rich. And at that point it was also true that we knew he would lose – but that was only true so long as it remained unsaid! (And maybe unthought as well.) Later, after the change in context, it was no longer true that we knew he would lose. At that point, it was also no longer true that we knew he would never be rich.

But wait. Don't you smell a rat? Haven't I, by my own lights, been saying what cannot be said? (Or whistled either.) If the story I told was true, how have I managed to tell it? In trendyspeak, is there not a problem of reflexivity? Does not my story deconstruct itself?

I said: S knows that P iff S's evidence eliminates every possibility in which not- P – Psst! – except for those possibilities that we are properly ignoring. That 'psst' marks an attempt to do the impossible – to mention that which remains unmentioned. I am sure you managed to make believe that I had succeeded. But I could not have done.

And I said that when we do epistemology, and we attend to the proper ignoring of possibilities, we make knowledge vanish. First we do know, then we do not. But I had been doing epistemology when I said that. The uneliminated possibilities were *not* being ignored – not just then. So by what right did I say even that we used to know?²⁴

In trying to thread a course between the rock of fallibilism and the whirlpool of scepticism, it may well seem as if I have fallen victim to both at once. For do I not say that there are all those uneliminated possibilities of error? Yet do I not claim that we know a lot? Yet do I not claim that knowledge is, by definition, infallible knowledge?

I did claim all three things. But not all at once! Or if I did claim them all at once, that was an expository shortcut, to be taken with a pinch of salt. To get my message across, I bent the rules. If I tried to whistle what cannot be said, what of it? I relied on the cardinal principle of pragmatics, which overrides every one of the rules I mentioned: interpret the message to make it make sense – to make it consistent, and sensible to say.

When you have context-dependence, ineffability can be trite and unmysterious. Hush! [moment of silence] I might have liked to say, just then, 'All of us are silent'. It was true. But I could not have said it truly, or whistled it either. For by saying it aloud, or by whistling, I would have rendered it false.

I could have said my say fair and square, bending no rules. It would have been tiresome, but it could have been done. The secret would have been to resort to 'semantic ascent'. I could have taken great care to distinguish between (1) the language I use when I talk about knowledge, or whatever, and (2) the second language that I use to talk about the semantic and pragmatic workings of the first language. If you want to hear my story told that way, you probably know enough to do the job for yourself. If you can, then my informal presentation has been good enough.

Notes

- 1 The suggestion that ascriptions of knowledge go false in the context of epistemology is to be found in Barry Stroud, 'Understanding Human Knowledge in General', in Marjorie Clay and Keith Lehrer (eds), *Knowledge and Skepticism* (Boulder: Westview Press, 1989); and in Stephen Hetherington, 'Lacking Knowledge and Justification by Theorising About Them' (lecture at the University of New South Wales, August 1992). Neither of them tells the story just as I do, however it may be that their versions do not conflict with mine.
 - 2 Unless, like some, we simply define 'justification' as 'whatever it takes to turn true opinion into knowledge' regardless of whether what it takes turns out to involve argument from supporting reasons.
 - 3 The problem of the lottery was introduced in Henry Kyburg, *Probability and the Logic of Rational Belief* (Middletown, CT: Wesleyan University Press, 1961), and in Carl Hempel, 'Deductive-Nomological vs. Statistical Explanation', in Herbert Feigl and Grover Maxwell (eds), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis: University of Minnesota Press, 1962). It has been much discussed since, as a problem both about knowledge and about our everyday, non-quantitative concept of belief.
 - 4 The case of testimony is less discussed than the others; but see C. A. J. Coady, *Testimony: A Philosophical Study* (Oxford: Clarendon Press, 1992), pp. 79–129.
 - 5 I follow Peter Unger, *Ignorance: A Case for Skepticism* (New York: Oxford University Press, 1975). But I shall not let him lead me into scepticism.
 - 6 See Robert Stalnaker, *Inquiry* (Cambridge, MA: MIT Press, 1984), pp. 59–99.
 - 7 See my 'Attitudes *De Dicto* and *De Se*', *The Philosophical Review* 88 (1979), pp. 513–43; and R. M. Chisholm, 'The Indirect Reflexive', in C. Diamond and J. Teichman (eds), *Intention and Intentionality: Essays in Honour of G. E. M. Anscombe* (Brighton: Harvester, 1979).
 - 8 Peter Unger, *Ignorance*, chapter II. I discuss the case, and briefly foreshadow the present paper, in my 'Scorekeeping in a Language Game', *Journal of Philosophical Logic* 8 (1979), pp. 339–59, esp. pp. 353–5.
 - 9 See Robert Stalnaker, 'Presuppositions', *Journal of Philosophical Logic* 2 (1973), pp. 447–57; and 'Pragmatic Presuppositions', in Milton Munitz and Peter Unger (eds), *Semantics and Philosophy* (New York: New York University Press, 1974). See also my 'Scorekeeping in a Language Game'.
 - 10 See Fred Dretske, 'Epistemic Operators', *The Journal of Philosophy* 67 (1970), pp. 1007–22, and 'The Pragmatic Dimension of Knowledge', *Philosophical Studies* 40 (1981), pp. 363–78; Alvin Goldman, 'Discrimination and Perceptual Knowledge', *The Journal of Philosophy* 73 (1976), pp. 771–91; G. C. Stine, 'Skepticism, Relevant Alternatives, and Deductive Closure', *Philosophical Studies* 29 (1976), pp. 249–61; and Stewart Cohen, 'How to be a Fallibilist', *Philosophical Perspectives* 2 (1988), pp. 91–123.
 - 11 Some of them, but only some, taken from the authors just cited.
 - 12 Instead of complicating the Rule of Belief as I have just done, I might equivalently have introduced a separate Rule of High Stakes saying that when error would be especially disastrous, few possibilities are properly ignored.
 - 13 A. D. Woozley, 'Knowing and Not Knowing', *Proceedings of the Aristotelian Society* 53 (1953), pp. 151–72; Colin Radford, 'Knowledge – By Examples', *Analysis* 27 (1966), pp. 1–11.
 - 14 See Edmund Gettier, 'Is Justified True Belief Knowledge?', this vol., ch. 7. Diagnoses have varied widely. The four examples below come from: (1) Keith Lehrer and Thomas Paxson Jr., 'Knowledge: Undeclared True Belief', *The Journal of Philosophy* 66 (1969), pp. 225–37; (2) Bertrand Russell, *Human Knowledge: Its Scope and Limits* (London: Allen and Unwin, 1948), p. 154; (3) Alvin Goldman, 'Discrimination and Perceptual Knowledge'; (4) Gilbert Harman, *Thought* (Princeton, NJ: Princeton University Press, 1973), p. 143.
- Though the lottery problem is another case of justified true belief without knowledge, it is not normally counted among the Gettier problems. It is interesting to find that it yields to the same remedy.
- 15 See Alvin Goldman, 'A Causal Theory of Knowing', *The Journal of Philosophy* 64 (1967), pp. 357–72; D. M. Armstrong, *Belief, Truth and Knowledge* (Cambridge: Cambridge University Press, 1973).
 - 16 See my 'Veridical Hallucination and Prosthetic Vision', *Australasian Journal of Philosophy* 58 (1980), pp. 239–49. John Bigelow has proposed to model knowledge-delivering processes generally on those found in vision.
 - 17 See Catherine Elgin, 'The Epistemic Efficacy of Stupidity', *Synthese* 74 (1988), pp. 297–311. The 'efficacy' takes many forms; some to do with knowledge (under various rival analyses), some to do with justified belief. See also Michael Williams, *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism* (Oxford: Blackwell, 1991), pp. 352–5, on the instability of knowledge under reflection.

- 18 Mixed cases are possible: Fred properly ignores the possibility W_1 which Ted eliminates; however Ted properly ignores the possibility W_2 which Fred eliminates. Ted has looked in all the desk drawers but not the file drawers, whereas Fred has checked the file drawers but not the desk. Fred's knowledge that Possum is not in the study is better in one way, Ted's is better in another.
- 19 To say truly that X is known, I must be properly ignoring any uneliminated possibilities in which not- X ; whereas to say truly that Y is better known than X , I must be attending to some such possibilities. So I cannot say both in a single context. If I say ' X is known, but Y is better known', the context changes in mid-sentence: some previously ignored possibilities must stop being ignored. That can happen easily. Saying it the other way around – ' Y is better known than X , but even X is known' – is harder, because we must suddenly start to ignore previously unignored possibilities. That cannot be done, really; but we could bend the rules and make believe we had done it, and no doubt we would be understood well enough. Saying ' X is flat, but Y is flatter' (that is, ' X has no bumps at all, but Y has even fewer or smaller bumps') is a parallel case. And again, ' Y is flatter, but even X is flat' sounds clearly worse – but not altogether hopeless.
- 20 Thanks here to Stephen Hetherington. While his own views about better and worse knowledge are situated within an analysis of knowledge quite unlike mine, they withstand transplantation.
- 21 A proof-theoretic version of this closure principle is common to all 'normal' modal logics: if the logic validates an inference from zero or more premises to a conclusion, then also it validates the inference obtained by prefixing the necessity operator to each premise and to the conclusion. Further, this rule is all we need to take us from classical sentential logic to the least normal modal logic. See Brian Chellas, *Modal Logic: An Introduction* (Cambridge: Cambridge University Press, 1980), p. 114.
- 22 See Dretske, 'Epistemic Operators'.
- 23 See Stalnaker, *Inquiry*, pp. 79–99.
- 24 Worse still: by what right can I even say that we used to be in a position to say truly that we knew? Then, we were in a context where we properly ignored certain uneliminated possibilities of error. Now, we are in a context where we no longer ignore them. If *now* I comment retrospectively upon the truth of what was said *then*, which context governs: the context now or the context then? I doubt there is any general answer, apart from the usual principle that we should interpret what is said so as to make the message make sense.

Contextualist Solutions to Epistemological Problems: Scepticism, Gettier, and the Lottery

Stewart Cohen

Among the many problems discussed in the epistemological literature, three that figure prominently are scepticism, the Gettier problem, and the lottery. In a recent paper, David Lewis proposes a theory of knowledge designed to solve all three problems.¹ Each, argues Lewis, can be handled by appealing to certain mechanisms of context-sensitivity – what he calls ‘rules of relevance’.

While others, myself among them, have proposed contextualist solutions to the problems of scepticism and the lottery, Lewis proposes to extend his contextualist approach to the Gettier problem.² I will argue that in so doing, Lewis’s contextualism overreaches – an appeal to context-sensitivity cannot solve the Gettier problem. The difference in this respect, between the Gettier problem, on the one hand, and scepticism and the lottery, on the other, will provide some insight into what motivates a contextualist treatment of an epistemological problem

1 Contextualism

While various kinds of epistemological theories have been called contextualist, I am here concerned with theories according to which the truth-value of a knowledge ascription is sensitive to certain facts about the speaker and hearers of the

context. Accordingly, for a particular subject *S*, and proposition *P*, one speaker could truly say ‘*S* knows *P*’ while at the same time another speaker in a different context truly says, ‘*S* does not know *P*’.

This way of viewing knowledge ascriptions is similar to a natural way of viewing flatness ascriptions.³ On this view, the truth-value of a flatness ascription is sensitive to context. For some particular *X*, one speaker could truly say ‘*X* is flat’ while at the same time, a speaker in another context could truly say ‘*X* is not flat’. For example, a group of (western) Coloradans may truly say that a particular road is flat while, at the same time in a different context, a group of Kansans with stricter standards may truly deny that the same road is flat.

We can think of the context-sensitivity of knowledge ascriptions in this way: For each context of ascription, there is a standard for how strong one’s epistemic position with respect to a proposition *P* must be in order for one to know *P*.⁴ Where two contexts differ with respect to this standard, a speaker in one context may truly say ‘*S* knows *P*’, while a speaker in the other truly says ‘*S* does not know *P*’. To complete the account, we need to specify how the standard for epistemic strength gets determined for each context. (More about this later.)

There are various ways of analysing this notion of the strength of one’s epistemic position. One could think of it as determined, at least in part, by the strength of one’s reasons or justification for believing *P*. On this view, the context-sensitivity of knowledge ascriptions derives from the context-sensitivity of standards for justification.

Originally published in *Australian Journal of Philosophy* 76, 2 (1998), pp. 289–308; reprinted by permission of Oxford University Press.

Consider again the analogy with flatness. We can think of a surface as being flat to varying degrees and we can also think of a surface as being flat *simpliciter*. What is the standard for how flat a surface must be to count as flat *simpliciter*? In different contexts, there can be different standards. Typically, when the topic of conversation is, e.g., roads, there will be a much stricter standard in contexts where the speaker and hearers are Kansans than when the speaker and hearers are Coloradans.

Analogously, we can think of a belief as being justified to various degrees and we can think of a belief as being justified *simpliciter*. On many views, being justified *simpliciter* is a necessary condition for a belief to be an instance of knowledge.⁵ What is the standard for how justified a belief must be to count as justified *simpliciter*? For the contextualist view, in different contexts, there can be different standards.

Because Lewis thinks justification is not a component of knowledge, he rejects this account of the context-sensitivity of knowledge ascriptions.⁶ To understand Lewis's account, consider another analogy with flatness ascriptions. We could view flatness ascriptions as involving a kind of implicit quantification: *X* is flat iff *X* has no bumps.⁷ We could then view the context as restricting the domain of quantification. In the Coloradans context, small hills do not count as bumps whereas in the Kansans context, they do.

For Lewis, knowledge ascriptions as well involve a kind of implicit quantification:

S knows *P* iff S's evidence eliminates every possibility in which not-*P* – Psst! – except for those possibilities that we are properly ignoring. (p. 554)⁸

Which possibilities can we properly ignore, i.e., what is the domain of 'every' in the definition? That is determined by the context. So on Lewis's view, the context determines how strong one's epistemic position must be with respect to *P*, i.e., the range of not-*P* possibilities one's evidence must eliminate, in order to know *P*.⁹

Which facts about the context determine which possibilities count – which possibilities must be eliminated and which can be properly ignored? Here Lewis provides a list of rules. In the language of relevant alternative theories, the rules tell us which alternatives are relevant – 'relevant, that is,

to what the subject does and doesn't know' (p. 554).

For our purposes, it is important to see that for some of the rules, what can be properly ignored depends on facts about the speaker and hearers of the context. As an example, we can consider the *Rule of Attention*. According to Lewis, if we are attending to a possibility, we are not properly ignoring it, for the simple reason that we are not ignoring it. So any possibility we are attending to is relevant in that very context. And as Lewis notes,

'We' [in the above definition] means: the speaker and hearers of a given context; that is, those of us who are discussing S's knowledge together. It is our ignorings not S's own ignorings, that matter to what we can truly say about S's knowledge. (p. 561)

So any possibility the speakers and hearers of a context are attending to is relevant in that very context.

Now just as there can be differences between what the speaker and hearers ignore, and what the subject ignores, so there can be differences between what the speaker and hearers of one context ignore and what the speaker and hearers of another context ignore. It follows that a possibility relevant in one context of ascription, may not be relevant in another. So on Lewis's view, a sentence ascribing knowledge to a particular subject at a particular time, can be true in the mouth of a speaker in one context and false in the mouth of another speaker in a distinct context.

II Scepticism

Contextualists argue that by appealing to context-sensitivity we can provide a satisfactory response to sceptical arguments. One strength of the contextualist approach is that it can account for the truth of our everyday knowledge ascriptions while still explaining the force of sceptical arguments. The basic idea is this: the sceptic's appeal to hypotheses (involving brains-in-a-vat, evil demons, etc.) creates a context where the standards for knowledge, i.e., the standards for how strong an epistemic position one must be in in order to know, are stricter than the standards that govern typical everyday contexts. In those 'sceptical' contexts, we fail to know anything. So

the contextualist concedes that there is some truth to scepticism. But a contextualist can limit the damage in a crucial way. For it remains true that we know many things in the typical everyday contexts where the standards are lower.¹⁰ In this way, the contextualist can explain the appeal of sceptical arguments while preserving the truth of our everyday knowledge ascriptions.

As we have seen, particular contextualist theories may differ with respect to how they conceive of the strength of one's epistemic position. They can also differ with respect to the mechanism that brings about the contextual shifts in the standards. On Lewis's version, the sceptic raises the standards for knowledge ascriptions by expanding the range of alternatives that must be eliminated. The sceptic brings this about by calling our attention to sceptical possibilities, thereby making them relevant (by the Rule of Attention). In those very contexts where we are attending to sceptical possibilities, we in fact fail to know many of the things we ordinarily take ourselves to know.

But sceptical possibilities are not relevant in every context. Most importantly, they are not relevant in everyday contexts, where we are not attending to sceptical possibilities. In those contexts, the standards for what must be eliminated are lower. So in these contexts, we can truly say of ourselves and others, that we know.

My aim, in this paper, is not to dispute Lewis's contextualist treatment of scepticism. I have defended a similar approach.¹¹ Rather I will take issue with Lewis's application of his contextualism to the Gettier problem, and his assimilation of the Gettier problem to the lottery problem. Lewis suggests that the lottery is a special case of the Gettier problem and argues that both can be solved by the same contextualist rule. Though I endorse a contextualist treatment of the lottery problem, I will argue that contextualism can shed no light on the Gettier problem. I argue that the lottery problem is of a piece with scepticism, not the Gettier problem.

III The Gettier Problem and the Lottery

The Gettier problem is (at least) to give a general account of why it is that certain cases of justified true belief fall short of knowledge. The lottery problem is to explain why it is (or perhaps merely seems) that no matter how great the number of

tickets in a lottery, i.e., no matter how great the odds are you will lose, you nonetheless fail to know you will lose.¹² We can sharpen the lottery problem by noting that, without knowing anything about the number of tickets, you can come to know you lose the lottery by reading the results of the drawing in the newspaper. This is puzzling because by simply increasing the number of tickets, we can make the odds of your losing conditional on the number of tickets, greater than the odds of your losing conditional on the newspaper report.

To solve these problems, Lewis appeals to two rules of relevance – the *Rule of Actuality* and the *Rule of Resemblance*. The Rule of Actuality states that 'the possibility that actually obtains is never properly ignored' (p. 554); The Rule of Resemblance states that if 'one possibility saliently resembles another [t]hen if one of them may not be properly ignored, neither may the other' (p. 556).

Now consider Lewis's application of these rules to a standard Gettier case:

The stopped clock is right twice a day. It says 4:39, as it has done for weeks. I look at it at 4:39; by luck I pick up a true belief. I've ignored the uneliminated possibility that I looked at it at 4:22 while it was stopped saying 4:39. That possibility was not properly ignored. It resembles actuality perfectly so far as the stopped clock goes. (p. 557)

By the Rule of Actuality, he cannot ignore the possibility that he looked, at 4:39, at the clock while it was stopped saying 4:39. And that possibility resembles – well enough – the possibility that he was looking at the clock at 4:22, while it was stopped saying 4:39. So, by the Rule of Resemblance, he cannot properly ignore this latter possibility either. Since this possibility is not eliminated by his evidence, he fails to know the time is 4:39.

Lewis wants to assimilate the lottery problem to the Gettier problem. He notes that though the lottery problem is a case of justified, true, belief that is not knowledge, it is not normally treated as a Gettier case. Yet, as he further notes, on his view it yields to the same treatment:

For every ticket, there is the possibility that it will win. These possibilities are saliently similar to one another; so either every one of them may

be properly ignored, or else none may. But one of them may not properly be ignored; the one that actually obtains. (p. 557)

So I cannot, by the Rules of Actuality and Resemblance, ignore the possibility, for any ticket, that it wins. In particular, I cannot ignore the possibility that my ticket wins. Since my evidence does not eliminate that possibility, I fail to know my ticket loses.

Is the lottery a kind of Gettier case? It is, of course, if by 'Gettier case' we mean 'a case of justified true belief that is not knowledge'. But the more interesting issue is whether Lewis is correct in claiming that the lottery yields to the same solution as the standard Gettier cases. I will argue that it does not. While the lottery problem is fundamentally a problem of context-sensitivity, the Gettier problem is not.

IV Two Kinds of Rules and the Status of the Rule of Resemblance

We can begin by examining a distinction between two kinds of rules for proper ignoring (rules of relevance) in Lewis's account. Some of Lewis's rules dictate that what can be properly ignored depends on facts about the speaker and hearers of the context. We saw this in the case of the Rule of Attention: what the speaker and hearers of the context are attending to affects what can be properly ignored.

But others of Lewis's rules are not like this. The Rule of Actuality says that the actual world is never properly ignored. But the rule does not refer to the actuality of the speaker of the context. Rather, Lewis insists, it is the actuality of the subject that is never properly ignored (p. 555). This distinction will matter only in cases where we are considering whether subjects know in worlds other than our own. But this is precisely what we are doing when we give our intuitive responses to merely possible cases, e.g., when we are considering the standard Gettier cases. So in those cases, the Rule of Actuality dictates that the subject's actuality cannot be ignored.

The Rule of Actuality does the work the truth condition does in traditional analyses of knowledge. It captures our intuition that you can not know *P*, if *P* is false. For our purposes, the important thing to note is that the operation of the Rule

of Actuality does not depend on who the speakers and hearers of the context are.

Let us call rules of relevance whose operation depends on facts about the speaker (and hearers) of the context, 'speaker-sensitive', and call rules of relevance whose operation depends on facts about the subject, 'subject-sensitive'. (As we shall see, a rule may be both subject-sensitive and speaker-sensitive.) Any theory that involves speaker-sensitive rules is contextualist in the sense I am concerned with.

Which kind of rule is the Rule of Resemblance? Consider its application to the lottery. Suppose we are considering whether *S* knows his ticket loses. The possibility that *S*'s ticket wins resembles actuality not because of anything pertaining to us, the speakers (and hearers) of the context. Rather the resemblance exists because of facts about *S*, the subject of the knowledge ascription. It depends on the fact that he holds a ticket in a fair lottery. This means that the Rule of Resemblance is subject-sensitive.

But there is more to the story. The Rule of Resemblance says that a possibility that *saliently* resembles actuality (or any possibility that is relevant by some rule other than Resemblance) cannot be properly ignored. The need for this qualification arises in connection with a variation of the lottery problem discussed by Lewis:¹³

Pity poor Bill! – He squanders all his spare cash on the pokies, the races, and the lottery. He'll be a wage slave all his days. We know he'll never be rich. (p. 565)

As Lewis suggests, intuitively, we know poor Bill will never get rich. But, intuitively, we do not know he loses the lottery. The problem is that his never getting rich, plus other things we know, entails that he loses the lottery. So if we do not know poor Bill loses the lottery, how can we know he'll never be rich?¹⁴

On Lewis's account, the Rules of Actuality and Resemblance determine that we fail to know poor Bill loses the lottery. 'But there is a loophole', according to Lewis,

... the resemblance was required to be salient. Salience as well as ignoring, may vary between contexts. Before, when I was explaining how the Rule of Resemblance applied to the lotteries, I saw to it that the resemblance between the many possibilities associated with the many tickets

was sufficiently salient. But this time, when we were busy pitying poor Bill for his habits and not for his luck, the resemblance of the many possibilities was not so salient. At that point, the possibility of Bill's winning was properly ignored; so then it was true to say that we knew he'd never be rich. Afterward, I switched the context. I mentioned the possibility that Bill might win, wherefore that possibility was no longer properly ignored... It was true at first that we knew that Bill would never be rich. And at that point it was also true that we knew he'd lose – but that was only true so long as it remained unsaid. Later after the change in context, it was no longer true that we knew he'd lose. At that point, it was also no longer true that we knew he'd never be rich. (pp. 565–6)

So in contexts where the lottery resemblances are not salient to us, we can properly ignore the possibility that poor Bill's ticket wins. And so in those contexts, we can know poor Bill will never get rich and that he will not win the lottery. But in contexts where the resemblances are salient, we can not properly ignore the possibility that Bill's ticket wins and so we do not know either of these things.¹⁵

When Lewis says the resemblances are required to be salient, does he mean 'salient to the subject of the ascription', or 'salient to the speaker ascribing knowledge'? In the case of poor Bill, this is a distinction without a difference. The way Lewis describes it, we are both ascribers and subjects. The issue for us is – What do we know about Bill? But when subject and speaker are distinct, we can see that Lewis must require that the resemblances be salient to the *speaker* (and hearers) of the context. Consider again how he applies the Rule of Resemblance to the Gettier cases. There the resemblances that undermine knowledge are not salient to the subject. Recall the subject S who by luck happens to be staring at a stopped clock at the very time, 4:39, displayed on the clock. According to Lewis, S fails to know the time because the possibility that he is looking at the clock at 4:22 while it is stopped saying 4:39, cannot be properly ignored. This possibility cannot be properly ignored because it resembles perfectly, as far as the stopped clock goes, the possibility that actually obtains. But this resemblance is not salient *for* S. If it were, this would not be a Gettier case since S would not even be justified in believing what the clock says. The resemblance is salient

for those of us who know S is in a Gettier situation, *viz.*, the speaker and hearers of the context. That is why we cannot ignore the possibility that the time is 4:22, with the clock stopped saying 4:39. And this explains why we cannot ascribe knowledge to S.

Because of the salience qualification, the Rule of Resemblance is speaker-sensitive. (We have seen that it is also subject-sensitive). This means that features of the context of ascription – facts concerning what resemblances are salient to the speaker (and hearers) – will determine which possibilities cannot, by this rule, be properly ignored. This aspect of the Rule of Resemblance, I shall argue, leads to a serious difficulty for Lewis's treatment of the Gettier problem.

V The Rule of Resemblance and Scepticism

Before returning to Lewis's treatment of the Gettier problem, I want to digress to show how the speaker-sensitivity of the Rule of Resemblance provides a way out of a problem Lewis raises for the application of this rule. 'We must apply the Rule of Resemblance with care', notes Lewis.

Actuality is a possibility uneliminated by the subject's evidence. Any other possibility *W* that is likewise uneliminated by the subject's evidence thereby resembles actuality in one salient respect: namely, in respect of the subject's evidence. That will be so even if *W* is in other respects very dissimilar to actuality – even if, for instance, it is a possibility in which the subject is radically deceived by a demon. Plainly, we dare not apply the Rules of Actuality and Resemblance to conclude that any such *W* is a relevant alternative – that would be capitulation to scepticism. (p. 556)

In response to this problem, Lewis makes an *ad hoc* stipulation that resemblances in respect of the subject's evidence alone do not count:

The Rule of Resemblance was never meant to apply to this resemblance! We seem to have an *ad hoc* exception to the Rule, though one that makes good sense in view of the function of attributions of knowledge. What would be better, though, would be to find a way to reformulate the Rule so as to get the needed exception

without *ad hocery*. I do not know how to do this. (pp. 556–7)

But there is a way for Lewis to avoid this sceptical problem without resorting to *ad hocery*. Possibilities that resemble actuality in respect of the subject's evidence are not, by the Rule of Resemblance, automatically relevant; they must *saliently* resemble actuality. That is the whole point of the Poor Bill case. But then scepticism does not threaten – that is, it does not threaten any more than a defender of context-sensitivity readily concedes. In normal, everyday contexts, where it is not salient that sceptical possibilities resemble actuality (in respect of the subject's evidence), those possibilities will not be relevant. Of course, the sceptic can make the resemblances salient to us and in that new context, the sceptical alternatives will be relevant. But this much the contextualist grants to the sceptic. This is how the contextualist explains the force of sceptical arguments. So in effect, Lewis is forced into his *ad hoc* restriction because he here treats the Rule of Resemblance as if it were merely subject-sensitive. But the rule's speaker-sensitivity enables us to avoid the threat of scepticism without resorting to *ad hocery*.

VI The Gettier Problem and Speaker-Sensitivity

Because the Rule of Resemblance is speaker-sensitive, it can both handle the Poor Bill variation of the lottery and avoid the threat of scepticism. But this very aspect of the rule undermines its applicability to the Gettier cases. Recall that according to Lewis's strategy for handling the Gettier cases, the subject fails to know *P* because there is an uneliminated not-*P* possibility that resembles actuality.¹⁶ But again, this by itself is not sufficient for the Rule of Resemblance to dictate that the not-*P* possibility cannot be properly ignored; the resemblance must be salient. The problem for Lewis is that there is nothing to guarantee that the resemblance will be salient. In some contexts the resemblance will be salient, but in others it will not. And in those contexts where the resemblance is not salient, the not-*P* possibility will be properly ignored and the subject will know.¹⁷

Surely this consequence of Lewis's theory is incorrect. To focus our intuitions, let's consider a particular Gettier case:

S sees what appears to be a sheep on the hill. But what S actually sees is a rock that looks, from that distance, to be a sheep. It happens though, that behind the rock, out of S's view, is a sheep.

In this case, the subject S has a justified true belief that there is a sheep on the hill, but S does not know there is a sheep on the hill.¹⁸ Why according to Lewis's account does S fail to know there is a sheep on the hill? Following Lewis's treatment of the stopped clock case, we can say that the possibility that there is no sheep on the hill but only a rock that looks like a sheep, resembles actuality. It resembles actuality perfectly with respect to the sheep-shaped rock. Thus by the Rules of Resemblance and Actuality, this possibility cannot be properly ignored.

Now consider A, standing next to S, who is unaware that S sees only a rock. The resemblance between the possibility that S sees a rock that looks like a sheep and actuality is not salient for A. A is not aware that S is in a Gettier situation of any kind. So according to Lewis's view, in A's context of ascription, the possibility that S sees merely a sheep-shaped rock can be properly ignored. Thus on Lewis's view, A truly ascribes knowledge to S. A can truly say 'S knows there is a sheep on the hill'.¹⁹

This strikes me as a strongly counterintuitive result. Surely it is very strange to suppose that there is any context of ascription in which one can truly say of S that he knows there is a sheep on the hill. The sentence, 'S knows there is a sheep on the hill' looks false (at that world and time), regardless of who happens to be uttering it. Consider again S's situation. S mistakenly thinks the rock he is seeing is a sheep. Surely S cannot in this way come to know there is a sheep on the hill, even if by luck, there happens to be a sheep on the hill hidden from view behind the rock. But again, if the resemblance relations that otherwise falsify ascriptions of knowledge are not salient in A's context, Lewis must hold that A's ascription of knowledge to S is true.²⁰

Does Lewis have a way to respond to this problem? One strategy he might adopt here is to bite the bullet and say that in A's context, A's ascription of knowledge to S is correct – and perhaps this would not be so difficult a bullet for him to bite. For as a contextualist, he can appeal to speaker-sensitivity to explain away the intuition that A cannot truly ascribe knowledge to S. He can hold

that our own context, at which knowledge cannot be truly ascribed to S, runs interference on our evaluation of what A says in his context. We confuse what can be properly ignored by us in our context of ascription with what can be properly ignored by A in his context of ascription. So the fact that we cannot truly say S knows there is a sheep on the hill prevents us from seeing that A can truly say S knows there is a sheep on the hill.

There would be nothing *ad hoc* about Lewis explaining away our intuition in this way. On the contrary, we could view it as a natural extension of the contextualist treatment of both the lottery and scepticism. A contextualist treatment of these problems must explain away certain of our intuitions as resulting from our confusing distinct contexts of ascription. The contextualist, while holding that our everyday knowledge ascriptions are correct, must acknowledge that when we are in the grips of a sceptical argument, those ascriptions seem mistaken. But, according to the contextualist, this intuition is misleading. It results from our confusing what we can truly say in the sceptical context we are in as we consider sceptical arguments, with what we can truly say in everyday contexts.²¹

And the same considerations apply to the Poor Bill variation of the lottery. We can truly ascribe knowledge that Bill loses the lottery (and that he will never get rich) in contexts where the resemblances that otherwise undermine knowledge are not salient. Nonetheless, in contexts where the resemblances are salient, it seems wrong to say that in other contexts where the resemblances are not salient, we can truly ascribe knowledge that Bill loses the lottery. Here again the contextualist must say that this intuition is misleading. It results from our confusing what can truly be said in our own context where the resemblances are salient, with what can truly be said in those contexts where the resemblances are not salient.

Now consider again our intuition that A's ascription to S of knowledge that there is a sheep on the hill is mistaken. In the very same way, Lewis can explain away our intuition as resulting from our confusing what can be truly ascribed in our own context where the relevant resemblance is salient with what can be truly ascribed in A's context, where the resemblance is not salient. This general strategy is central to the general contextualist view he is defending.

Of course, some may find contextualism to be implausible in general. For those who take this

view, the application of this strategy to the Gettier case would be just another example of the kind of implausible line the theory takes in the cases of scepticism and the lottery. But I, as a proponent of a contextualist treatment of scepticism and the lottery, do not want to endorse this position. Still, I find it very implausible that we can truly ascribe knowledge in the Gettier case in *any* context. Thus, I will argue that this kind of contextualist approach, even if correct in the cases of scepticism and the lottery, cannot be extended to the Gettier problem.

VII The Scope of Speaker-Sensitivity

On Lewis's view, the Rule of Resemblance is central to solving the Gettier problem. We can think of this rule as corresponding, roughly, to the Gettier or fourth condition in a standard analysis of knowledge. As we noted earlier, the Rule of Actuality corresponds to the truth condition.

Now consider Lewis's *Rule of Belief*:

A possibility that the subject believes to obtain is not properly ignored, whether or not he is right to so believe. *Neither is one that he ought to believe to obtain – one that evidence and arguments justify him in believing – whether or not he does so believe.* [my emphasis] (p. 555)

The second part of this rule corresponds roughly to some of the phenomena that motivate the justification condition in a standard analysis of knowledge.

So each of the Rules of Actuality, Belief, and Resemblance corresponds to some element in a traditional analysis of knowledge. But on Lewis's view, there is an asymmetry in the way these rules are applied. This asymmetry results from the fact that of the three, only the Rule of Resemblance is speaker-sensitive. Note that the Rule of Belief, like the Rule of Actuality, is only subject-sensitive. According to this rule, we cannot ignore a possibility the subject believes, or merely ought to believe, to obtain.

Compare three cases. S₁ is looking at a sheep-shaped rock that happens to have a sheep behind it. S₂ is actually looking at a sheep under normal conditions, though he has evidence that justifies him in believing he is in fact looking at a sheep-shaped rock on a sheepless hill. S₃ is in fact looking at sheep-shaped rock on a sheepless hill. All three

subjects believe there is a sheep on the hill. But none of these subjects knows there is a sheep on the hill. From the standpoint of traditional analyses of knowledge, S_1 fails to know because though he has a justified true belief that there is a sheep on the hill, he is in a Gettier situation. S_2 fails to know because his evidence justifies him in believing that, despite the appearances, there is no sheep on the hill, making his belief that there is a sheep on the hill unjustified. And S_3 fails to know because his belief that there is a sheep on the hill is false.

On Lewis's view, in each case, the failure of the subject to know there is a sheep on the hill is explained by the rules of relevance. We cannot truly ascribe knowledge to any of them, because in each case we may not properly ignore the possibility that there is only a sheep-shaped rock on the hill. S_1 is in a situation where that possibility saliently (for us) resembles actuality. So by the Rules of Resemblance and Actuality, we cannot properly ignore it. S_2 ought to believe that possibility obtains. So by the Rule of Belief, we cannot properly ignore it. And for S_3 , that possibility is actual. So, by the Rule of Actuality, we cannot properly ignore it.

The rules of relevance deliver the same result for what we, in our context, can truly say about each subject: In each case, we are correct in denying that the subject knows there is a sheep on the hill. But suppose we consider whether it is correct to ascribe knowledge to these subjects in certain contexts of ascription other than our own. Consider speaker A who is not aware that S_1 is in a Gettier situation; A thinks S_1 is, in fact, looking at a sheep. Nor is A aware that S_2 has evidence that justifies him in believing that there is no sheep on the hill. Nor is A aware that S_3 merely seems to see a sheep. Given only what is *salient* in A's context, all three subjects appear to be in the same epistemic situation, *viz.*, looking at a sheep under normal conditions. This leads A to say all three subjects know there is a sheep on the hill.

Are A's ascriptions of knowledge to S_1 , S_2 , and S_3 , true? (We can assume hearers with the same information as A.) Consider first A's ascription of knowledge to S_1 . Though the possibility that S_1 sees a rock with no sheep behind it resembles actuality (S_1 's actuality), that resemblance is not salient for A. So, as we noted in section VI, as far as the Rule of Resemblance goes, A can properly ignore that possibility. Thus A can truly ascribe knowledge to S_1 .

But matters are different for A's ascriptions to S_2 and S_3 . S_2 ought to believe he sees a rock that looks like a sheep. So by the Rule of Belief, A may not properly ignore that possibility. Though it is not salient in A's context that S_2 ought to believe he sees merely a rock, the Rule of Belief does not require that it be salient. Thus A cannot truly ascribe knowledge to S_2 .

Analogously, in the case of S_3 . A may not ignore the possibility that S_3 sees merely a sheep-shaped rock. For that possibility is actual and so by the Rule of Actuality. A cannot properly ignore it. Though it is not salient in A's context that this possibility is actual, the Rule of Actuality does not require that it be salient. Thus A cannot truly ascribe knowledge to S_3 .

Why should there be this asymmetry between knowledge ascriptions to S_1 , on the one hand, and knowledge ascriptions to S_2 and S_3 , on the other? The asymmetry results from the fact that the Rule of Resemblance is speaker-sensitive whereas the rules of Belief and Actuality are not. Because of this, the truth-value of knowledge ascriptions to S_1 can vary with the speaker whereas the truth-value of knowledge ascriptions to S_2 and S_3 cannot. This seems right for S_2 and S_3 . Intuitively, neither S_2 nor S_3 knows there is a sheep on the hill, regardless of what is salient to the speaker (and hearers) of the context – S_2 , because his evidence justifies him in believing there is no sheep on the hill and S_3 because his belief that there is a sheep on the hill is false. Our intuitions here provide a basis for thinking it is part of the fixed (across contexts) truth conditions for 'S knows P ' that P is true and that S's evidence does not justify him in believing not- P . If so, then we have a rationale for holding that knowledge ascriptions to S_2 and S_3 are false (at those worlds and times), regardless of who is making the ascription. But analogously, our intuitions provide the same kind of rationale for taking it as part of the fixed truth conditions for 'S knows P ' that S is not in the kind of Gettier situation S_1 is in. Intuitively, because of his situation, S_1 fails to know he sees a sheep, regardless of what is salient to the speaker of the context. Thus we have equally good reason to hold that, contrary to what Lewis's theory entails, knowledge ascriptions to S_1 are false, regardless of who is making them. This suggests that in order for the Rule of Resemblance to solve the Gettier problem, it should be construed as speaker-*insensitive*. This would require that the salience qualification be eliminated.²²

Of course an asymmetry can be eliminated from either direction. Suppose we do have as good a rationale for holding that S_1 fails to know, irrespective of the speaker, as we do for holding that S_2 and S_3 fail to know, irrespective of the speaker. We could still reject that rationale for all three cases. After all, we saw in section IV how Lewis could employ a contextualist strategy to reject the rationale in the case of S_1 . This strategy seeks to explain away the intuition that S_1 fails to know in contexts like A 's as resulting from a confusion of contexts. We confuse contexts like A 's, where the resemblance between actuality and the possibility at which there is only a sheep-shaped rock, is not salient, with contexts like our own where this resemblance is salient. Since in our context, S_1 fails to know, we mistakenly think S_1 fails to know in A 's context as well.

Now we could, in the same way, seek to explain away our intuition that S_2 and S_3 fail to know regardless of the speaker. We could do this by construing the Rules of Belief and Actuality as speaker-sensitive, on the model of the Rule of Resemblance. Viewed in this way, the Rule of Belief would say that a possibility that the subject *saliently* believes, or saliently ought to believe, to obtain cannot be properly ignored. And the Rule of Actuality would say that when a possibility is *saliently* actual, it cannot be properly ignored. On this view, when we consider whether S_2 knows, we cannot properly ignore the possibility that there is only a sheep-shaped rock on the hill. In our context, it is salient that S_2 ought to believe this possibility obtains. Nor can we ignore this possibility when we consider whether S_3 knows. In our context, it is salient that this possibility is actual (for S_3). But in A 's context, neither of these facts is salient. So, in each case, A can properly ignore this possibility. So construing the rules in this way would allow that A could truly say that each of S_2 and S_3 knows there is a sheep on the hill. Our intuition that A , in his context, falsely ascribes knowledge to both S_2 and S_3 results from our confusing our own context with A 's. This strategy strikes me as no less plausible in the cases of S_2 and S_3 than in the case of S_1 .

So there are two options for eliminating the asymmetry between the way Lewis's theory handles the case of S_1 and the way his theory handles the cases of S_2 and S_3 . We could eliminate the salience qualification from the Rule of Resemblance thereby making it speaker-insensitive. Or we could add a salience qualification to the Rules

of Belief and Actuality thereby making them speaker-sensitive. Is there any reason to prefer one option to the other? I think that the best explanation for our intuition that A is mistaken in ascribing knowledge to all three subjects is that *even relative to A 's context*, each of them in fact fails to know – S_1 because he is looking at a sheep-shaped rock with a sheep behind it, S_2 because his evidence justifies him in believing there is no sheep, and S_3 because he is looking at a hill with no sheep on it. A *mistakenly* ascribes knowledge to each of them for the simple reason that he is unaware, in each case, of the knowledge-defeating circumstance. So, I would hold that none of the three rules is speaker-sensitive, that all three subjects, S_{1-3} , fail to know there is a sheep on the hill, regardless of what is salient to the speaker.

But how do I reconcile taking this view for these cases with my defence of the contextualist approach to scepticism (and the Poor Bill variation of the lottery)? If what I say is the correct explanation for our intuition that A is mistaken in ascribing knowledge to S_1 , S_2 , and S_3 , why should we not say the same thing about our sceptical intuitions toward our everyday knowledge ascriptions? As we have noted, when in the grips of a sceptical argument, we feel the strong intuitive pull of saying that our knowledge ascriptions are mistaken, even those we make in everyday contexts. But here the contextualist denies that our sceptical intuitions indicate that our everyday knowledge ascriptions are mistaken – that sceptical possibilities are relevant even in contexts where they are not salient to us. Rather the contextualist appeals to speaker-sensitivity to explain away those intuitions as resulting from a confusion of contexts.

So on what basis do I claim that we should invoke speaker-sensitivity to explain away our sceptical intuitions toward our everyday knowledge ascriptions, but not our 'sceptical' intuitions toward ascriptions of knowledge to S_1 , S_2 , and S_3 ? If we take these latter intuitions at face value, as indicating that A 's ascriptions of knowledge to S_1 , S_2 , and S_3 are mistaken even in A 's context, why not take our sceptical intuitions at face value, as indicating that our everyday knowledge ascriptions are false, even in everyday contexts? This would mean viewing sceptical possibilities as relevant to any knowledge ascription regardless of what is salient to the speaker. Then, just as we explain A 's mistaken ascriptions of knowledge to S_{1-3} as resulting from A 's ignorance of certain features of the subject's epistemic situation, so the sceptic

could explain why we have been mistakenly ascribing knowledge to ourselves in everyday contexts as resulting from our ignorance (prior to our initiation to sceptical arguments) of certain facts concerning our epistemic situation, *viz.*, that our evidence does not eliminate sceptical possibilities.

As it turns out, there is an important difference between our sceptical intuitions and our intuitions about S_{1-3} – a difference that provides a rationale for appealing to speaker-sensitivity in the explanation of our sceptical intuitions, but not in the explanation of our intuitions toward knowledge ascriptions to S_{1-3} . Consider first our sceptical intuitions. Strong as they sometimes may be, they conflict with other strong intuitions we have, *viz.*, our common sense intuitions concerning what we know. As Lewis notes:

We have all sorts of everyday knowledge, and we have it in abundance. To doubt that would be absurd. At any rate, to doubt in any serious and lasting way would be absurd; and even philosophical and temporary doubt, under the influence of argument, is more than a little peculiar. (p. 549)

Of course, Lewis does not mean to be denying that there is any problem of scepticism. He acknowledges that, when considering sceptical arguments, we often feel a strong pull toward saying that our knowledge ascriptions are false. His theory is designed, in part, to explain these intuitions. It's just that these sceptical intuitions are not stable. We also find, while still engaged in philosophical reflection, that we have a tendency to shift to a perspective from which that conclusion is difficult to accept – 'How could it be that I fail to know I have a hand? ... Surely I know that!' But then I can again be overcome by the pull of the sceptical arguments and begin to doubt that I do know anything – even that I have hands. But this time around, the sceptical intuition is no more stable than before. This kind of vacillation is a fairly robust feature of our intuitions about scepticism. When we think about scepticism, we find ourselves pulled in inconsistent directions – we find ourselves shifting back and forth between thinking we fail to know and thinking that this conclusion is absurd.

This contrasts starkly with our intuitions toward ascriptions of knowledge to S_{1-3} . S_1 is looking at a sheep-shaped rock that happens to have a sheep behind it. Here we have a strong

and stable intuition that S_1 does not know there is a sheep on the hill. In no way does it seem 'absurd' or 'more than a little peculiar' to deny that S_1 knows there is a sheep on the hill. We deny unequivocally that S_1 knows. Even if previously, because of our ignorance of S_1 's situation, we were inclined to say he knows there is a sheep on the hill, once we learn that he actually sees only a rock, we would not be inclined, in the least, to say he knows. We would not find ourselves vacillating between saying that S_1 does not know and saying that he does.

Similarly we have a strong and stable intuition about ascriptions of knowledge to S_2 and S_3 . Intuitively each fails to know there is a sheep on the hill, S_2 because his evidence justifies him in believing there is no sheep on the hill, and S_3 he is looking at a sheepless hill. In neither case do we find our intuitions pulled in opposing directions.²³

But again, this kind of stability is lacking in the intuitions we have in contexts where we are considering sceptical arguments; these intuitions are notoriously unstable. Thus while nothing stands in the way of taking our intuitions toward S_{1-3} at face value, taking our sceptical intuitions at face value is problematic.

Here then is the opening for the contextualist to explain away our sceptical intuitions toward our everyday knowledge ascriptions by appealing to speaker-sensitivity. According to the contextualist, when we are in a sceptical frame of mind, we are in a context where sceptical possibilities are relevant. But again we are of two minds. We can also find it compelling that we know many things. This is because even though in our own context, sceptical possibilities are relevant, we can still evaluate knowledge ascriptions relative to other contexts where sceptical possibilities are not relevant. So even when a sentence ascribing knowledge is false in our own mouths, we can still evaluate it as true in the mouths of others (or our own mouths in other contexts). Thus, the instability of our intuitions toward our everyday knowledge ascriptions results from our alternatively evaluating them relative to our current sceptical contexts and other non-sceptical contexts.²⁴

A similar phenomenon occurs in the Poor Bill version of the lottery. It seems intuitive to say we know he will never get rich. Yet we also find it intuitive that we do not know he will lose the lottery. But these intuitions take us in opposing directions. We know he'll never get rich only if we know he'll lose the lottery. Thus we find ourselves

vacillating between thinking we know he'll never get rich and so that he'll lose the lottery, and thinking we know neither of these things. The contextualist holds that our opposing intuitions result from our evaluating these knowledge ascriptions relative to different contexts. In some contexts, we know he'll never get rich (and that he'll lose the lottery). In others, we fail to know that he'll never get rich (and that he'll lose the lottery).

Of course an appeal to speaker-sensitivity is not the only way to explain the instability of our intuitions. There are various sceptical explanations, as well, e.g., ones that appeal to force of habit. Which explanation is the best is not an issue we need confront in this paper.²⁵ The point is that an appeal to speaker-sensitivity has at least some initial plausibility as an explanation for why our intuitions are unstable. This provides a motivation for appealing to speaker-sensitivity in our treatment of scepticism and the lottery. Where no such instability exists, as with our intuitions toward A's knowledge ascriptions to S_{1-3} , this motivation for appealing to speaker-sensitivity is absent. In these cases there is nothing for an appeal to speaker-sensitivity to explain. This makes it much more plausible to take at face-value our strong intuition that A's ascriptions of knowledge to S_{1-3} are simply mistaken.

VIII The Gettier Problem, the Lottery, and the Rule of Resemblance

I have been arguing that it is considerably more plausible to appeal to speaker-sensitivity for knowledge ascriptions when our intuitions regarding them are unstable. As it turns out, Lewis's analysis is consonant with this view insofar as scepticism, the Poor Bill version of the lottery, and A's ascriptions of knowledge to S_2 and S_3 are concerned. In the cases of scepticism and poor Bill, where our intuitions are unstable, the operative rules of relevance – the Rule of Attention and the Rule of Resemblance – are speaker-sensitive. This allows the truth-value of knowledge ascriptions to vary in ways that explain the instability of our intuitions. In the cases of S_2 and S_3 , where our intuitions are stable, the operative rules of relevance – the Rule of Belief and the Rule of Actuality – are not speaker-sensitive. This prevents the truth-value of the knowledge ascription from shifting with the context, thus explaining the stability of our intuitions.

But as we have seen, Lewis's analysis diverges from this view in the (Gettier) case of S_1 . Here, even though our intuitions are stable, the operative rule of relevance – the Rule of Resemblance – is speaker-sensitive. If the argument of the previous section is correct, Lewis is mistaken in trying to handle this case with a speaker-sensitive rule. More generally, it is problematic to use a speaker-sensitive rule of relevance to solve the Gettier problem. There is no reason not to view the subject's failure to know in these Gettier cases as fixed across contexts of ascription – as holding regardless of who the speaker is.

So any rule that solves the Gettier problem must be speaker-insensitive. Thus, for the Rule of Resemblance to solve the Gettier problem, the salience qualification must be eliminated. But this would raise two problems. First, without the salience qualification, the sceptical consequences of the rule return. Recall that sceptical possibilities resemble actuality perfectly with respect to the subject's evidence. So, as Lewis notes, the Rule of Resemblance would seem to entail that sceptical possibilities are relevant, i.e., that they cannot be properly ignored. I argued in section V that, because of the salience qualification, the Rule of Resemblance can avoid this consequence. For in everyday contexts, the resemblances between sceptical possibilities and actuality are not salient. Without the salience qualification, this response is no longer available.

Lewis's own response to this difficulty was to make an *ad hoc* stipulation that these resemblances do not count. But this is really to concede that the rule does not solve the problem. Solving the Gettier problem requires either further revision of the Rule of Resemblance or perhaps a different rule altogether.

A second problem for eliminating the salience qualification from the Rule of Resemblance is that doing so would render the rule unable to handle the poor Bill variation of the lottery problem. For it would rule out our knowing of someone that he loses the lottery, in any context of ascription. But then it rules out our knowing, in any context of ascription, anything that entails of someone that he loses the lottery.²⁶ But it does seem that we can know many things that entail that a certain person loses the lottery. For example, we can know, in some contexts anyway, that poor Bill will never get rich.²⁷ The whole point of the salience qualification was to make the rule speaker-sensitive thereby

allowing knowledge in some contexts that a certain person loses the lottery.

These two problems for the unqualified Rule of Resemblance show that resemblance is too pervasive a phenomenon to be appealed to in an unrestricted way. As we have seen, a natural move is to invoke speaker-sensitivity and restrict the rule by salience. But this very feature of the rule makes it incapable of solving the Gettier problem.²⁸

We can now see that there is good reason to reject Lewis's assimilation of the lottery to the Gettier problem. Lewis notes that though the lottery seems to be a case of justified, true, belief that is not knowledge, it is not normally counted as a

Gettier case. But according to Lewis, it can be solved by the same rule he invokes to handle the traditional Gettier cases, *viz.*, the Rule of Resemblance. We have just seen that this claim is doubtful. The lottery (given the poor Bill variation) requires speaker-sensitivity whereas the Gettier problem requires speaker-*ins*sensitivity. The contextualist, as we have seen, invokes speaker-sensitivity to solve the problem of scepticism. The fact that an adequate treatment of the lottery requires an appeal to speaker-sensitivity as well suggests that the lottery has more in common with scepticism than with the Gettier problem.

Notes

- 1 [10] All Lewis page references are to this paper.
- 2 For other contextualist accounts, see Unger [12], [13], Cohen [2], [3], [4], and DeRose [5].
- 3 See [9], [13], [6]. I would argue for this kind of contextualism for many predicates, e.g., 'happy', 'tall', 'old', 'rich' . . .
- 4 I argue for such a view in [2]. There I talk about how strong our epistemic position must be with respect to alternatives to P in order for us to know P. Also see DeRose [5].
- 5 So for a belief to be justified *simpliciter*, it must be justified to the minimum degree necessary for knowledge. I do not claim that this is the only notion of justification *simpliciter*.
- 6 Lewis does not subscribe to this account of context-sensitivity because he holds that justification is neither necessary nor sufficient for a true belief to be knowledge. But I do not find his reasons for this view to be very convincing.

Justification is not necessary for knowledge, according to Lewis, because no 'argument supports our reliance on perception, on memory, and on testimony. Yet we do gain knowledge by these means' (p. 551). But most accounts of justification for perception, memory, and testimony do not require our having anything like a supporting argument. Here, Lewis seems to rely on an overly restrictive conception of justification.

Justification is not sufficient for a true belief to be knowledge, according to Lewis, because in a lottery, one does not know one loses regardless of the number of tickets. Yet the greater the number of tickets, the greater one's justification (p. 551). This argument is somewhat puzzling. Since Gettier, few epistemologists think of justification as sufficient for a true belief to be knowledge. And Lewis suggests that the lottery is a kind of Gettier case. (I argue against this in section VIII). More importantly, Lewis ultimately

allows that one can know, in certain contexts, that one loses the lottery. But then I see no reason why we should not say that in those contexts, one's justification is sufficient for one to know.

As we shall see, for Lewis, knowing P is a matter of one's evidence eliminating some alternatives to P and one's being able to properly ignore the rest of the alternatives. Certainly one's evidence eliminating alternatives is a component of justification as it is construed by many epistemologists. Moreover, whether or not one can properly ignore certain alternatives turns out to depend on one's evidence. This is true for the Rule of Resemblance and the Rule of Belief. So, though he denies it, I think it is fair to say Lewis's analysis of knowledge involves standards for justification – as many epistemologists think of justification, anyway.

- 7 Dretske [6], Unger [13].
- 8 Lewis says, '... a possibility W is uneliminated iff the subject's perceptual experience and memory in W exactly match his perceptual experience and memory in actuality' (p. 553).
- 9 On Lewis's view, the extent to which one's evidence eliminates possibilities is only part of what constitutes the strength of one's epistemic position. Some of the conditions that determine whether or not one can properly ignore a not-P possibility can be viewed as contributing to the strength of one's epistemic position, as well. This holds, in particular, for the rule of resemblance and the rule of belief.
- 10 For stylistic reasons, following Lewis, I will not always be careful about formulating the contextualist thesis metalinguistically. So instead of saying that a sentence containing the knowledge predicate can be true in one context and false in the other, I will say that whether we know can vary across contexts. Strictly speaking, though, the metalinguistic formulation should be used.

- 11 See [2]. On my version of contextualism, the standard governs justification and the mechanism is a Rule of Salience which is very similar, and perhaps equivalent to, Lewis's Rule of Attention. On DeRose's version in [5], the standard governs what he calls 'truth tracking' and the mechanism is the Rule of Sensitivity.
- 12 There are, of course, other problems involving lotteries, see [10].
- 13 Gilbert Harman discusses a case like this in [7].
- 14 Lewis endorses the principle that knowledge is closed under strict implication (p. 564). A weaker principle says that knowledge is closed under known implication. Either way, the problem of poor Bill arises. I defend the closure principle in [2].
- 15 In Lewis's discussion, he notes that the possibility that poor Bill wins is also made relevant by the Rule of Attention, since Lewis mentions that possibility.
- 16 In some Gettier cases, the explanation appeals to the Rule of Resemblance working in conjunction with some other rule. See p. 557.
- 17 The possibility will be properly ignored, at least so far as the Rule of Resemblance goes. And Lewis says the Rule of Resemblance explains why the subject fails to know in a Gettier situation. Moreover, none of Lewis's other rules seem applicable.
- 18 This example is taken from Chisholm [1]. Nothing I say hinges on using this case rather than the stopped clock case. I use the former only because, by my lights, the intuition that the subject fails to know in this case is even more vivid than in the latter.
- 19 We might be able to construe salience in a way that results in the resemblance being salient. Let's say that the resemblance is strongly salient if it is salient that the possibility resembles actuality, and the resemblance is weakly salient if the features in virtue of which the possibility resembles actuality are salient. A possibility can be weakly salient without being strongly salient. Consider the feature in virtue of which the possibility that S sees a sheep-shaped rock with no sheep behind it resembles actuality, *viz.*, the sheep-shaped rock. If A himself is looking at the sheep-shaped rock (thinking that it is a sheep), then the resemblance is weakly salient for A but not strongly salient. But of course, nothing guarantees that in Gettier cases, the relevant resemblances will be even weakly salient. A may not be looking at the sheep-shaped rock.
- As another example, consider the bogus barn case. The possibility that S sees a bogus barn resembles actuality in virtue of the abundance of bogus barns in the vicinity of the actual barn. But this need not be salient to someone ascribing knowledge to S. Moreover, if we formulate the rule of resemblance in terms of weak salience, then the sceptical implications of the rule return. For the features in virtue of which, e.g., the brain-in-a-vat hypothesis resembles actuality are salient in everyday contexts. So these possibilities will (weakly) saliently resemble actuality in those contexts.
- 20 Given what he says in footnote 24, Lewis would hold that there is no general answer to whether we can say in the object language that A is in a position to truly say S knows. Of course we can avoid whatever problem there may be here if we, as Lewis suggests at the end of the paper, ascend semantically, i.e., describe the case metalinguistically.
- 21 I take this line in [2], [3], and [4].
- 22 We would then need a different rule to handle the Poor Bill version of the lottery. More about this in section VIII.
- 23 It may be that if we later forget that e.g., S₃ is looking at a sheepless hill, we will then find it intuitive that he knows. This is also true of scepticism – if we forget about sceptical possibilities we will find it intuitive that we know. But once it is pointed out to us that S₃ is seeing only a sheep-shaped rock, we will again think, unequivocally, that we made a mistake. Our sceptical intuitions are not like that.
- 24 It also may be that our vacillations toward our knowledge ascriptions indicate that through subtle changes in ourselves as speakers and hearers, changes in our intentions, focus, purposes, etc. we bring about shifts in the set of relevant possibilities, i.e., we ourselves shift contexts. Sometimes we can indicate such a shift in our intentions and purposes by speaking in a certain tone of voice: 'C'mon, you know you're not a brain-in-a-vat!' If this is correct, we would have to treat the rule of attention as defeasible. See Lewis's discussion of bending the rules of co-operative conversation, p. 560.
- 25 I defend the contextualist explanation over sceptical explanations in [4].
- 26 This assumes the deductive closure principle. See my footnote 14.
- 27 An example given by Harman in [7] is perhaps even more compelling. We can know S will be in New York tomorrow, even though if he wins the lottery he will be in New Jersey instead, collecting his winnings, i.e., even though knowing S will be in New York involves knowing S will lose the lottery.
- 28 I argue in [2] that both scepticism and the lottery can be handled by appealing to the same rule – a rule of salience. Though space considerations preclude a full discussion here, it may be that even on Lewis's view both problems can be solved by the same rule. According to Lewis, both problems result from speaker-sensitivity. And though on Lewis's view, both are handled by speaker-sensitive rules, each is handled by a different rule – scepticism, by the Rule of Attention, and the lottery, by the Rule of Resemblance. But it is not clear why Lewis needs to appeal to the Rule of Resemblance to solve the lottery problem. For it looks as if any possibility relevant by the Rule of Resemblance will also be relevant by the Rule of Attention.

Recall that for a possibility W to become relevant by the Rule of Resemblance, two conditions must be met. First, W must resemble actuality (or some other possibility relevant by some other rule). And second, the resemblance between W and actuality must be salient in the context. But if the resemblance between W and actuality is salient in the context, then W itself is salient in that context. For example, because it is a fair lottery, the possibility that my ticket wins (as well as the possibility for every other losing ticket, that it wins) resembles the possibility that (the winning ticket) T wins with respect to the set-up of the lottery. Now if it is salient that the possibility that my ticket wins resembles the possibility that T wins, then trivially the possibility that my ticket wins (along with the pos-

sibility that T wins) is salient as well. More generally, if possibilities saliently resemble one another in a context, then those possibilities themselves are salient in that context.

Does it follow that if possibilities saliently resemble one another in a context, then those possibilities are being attended to in that context? What exactly is the relationship between salience and attention? These issues are too large to explore here. Suffice it to say that for Lewis, salience is a kind of psychological notion closely related to attention. Though something can be salient without actually being attended to – think of the cartoon character, Mr Magoo – it is not clear why this difference should matter for whether an alternative is relevant in a context.

References

1. Roderick Chisholm, *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).
2. Stewart Cohen, 'How to be a Fallibilist', *Philosophical Perspectives* 2 (1988), pp. 581–605.
3. Stewart Cohen, 'Skepticism, Relevance, and Relativity', in *Dretske and his Critics* (Oxford: Basil Blackwell, 1991), pp. 17–37.
4. Stewart Cohen, 'Skepticism and Everyday Knowledge Attributions', in *Doubting* (Dordrecht: Kluwer, 1990), pp. 161–70.
5. Keith DeRose, 'Solving the Sceptical Problem', this vol., ch. 38.
6. Fred Dretske, 'The Pragmatic Dimension of Knowledge', *Philosophical Studies* 40 (1981), pp. 363–78.
7. Gilbert Harman, *Thought* (Princeton: Princeton University Press, 1974).
8. Henry Kyburg, 'Conjunctivitis', in *Induction, Acceptance, and Rational Belief* (Dordrecht: Reidel, 1970), pp. 55–82.
9. David Lewis, 'Scorekeeping in a Language Game', *Journal of Philosophical Logic* 8 (1979), pp. 339–59.
10. David Lewis, 'Elusive Knowledge', this vol., ch. 39.
11. Gail Stine, 'Skepticism, Relevant Alternatives, and Deductive Closure', *Philosophical Studies* 29 (1976), pp. 249–61.
12. Peter Unger, 'The Cone Model of Knowledge', *Philosophical Topics* 14 (1986), pp. 125–78.
13. Peter Unger, *Philosophical Relativity* (Minneapolis: University of Minnesota Press, 1984).

PART X

Relativism

Introduction

Each of the authors in this section argues that attributions of justification in some way lack objectivity. Each then draws negative conclusions from this about the value of pursuing epistemology as it has been traditionally conceived in analytic philosophy.

Michael Williams argues, persuasively, that if there is such a thing as *knowledge of the external world*, the kind of knowledge the Cartesian skeptic questions, it seems impossible for us to see ourselves as having it. That is, the skeptic would carry the day. But he asks: *is there such a thing as knowledge of the world?* His answer is no. The concept of knowledge of the external world is a theoretical concept, unlike practical concepts such as the concept of a chair, and so it lacks application entirely unless there is an appropriate unified domain of reality whose contours are there for it to match. There is no such epistemic domain. There could be only if (empirical) beliefs divided into two classes: those that could only be known on the basis of beliefs about immediate experience, i.e., beliefs about the external world, and those that could be known directly from immediate experience. Yet an examination of our practices in attributing knowledge and justification testifies to the conclusion that beliefs do not divide into these epistemic categories, nor in fact into any objective epistemic categories. Truths about what justifies what vary from context to context depending on the accepted aims and enterprises constitutive of the context. Thus, in the context of geology, challenging the justification of beliefs about whether there really are rocks is misplaced or perhaps absurd.

Williams describes his view as a form of contextualism. But it is a contextualism quite different

than the contextualisms appearing in the previous section. The contextualist theories of DeRose, and to a lesser extent Lewis, acknowledge substantial objective characteristics as context-invariant partial determinants of the truth of knowledge attributions. For DeRose, there are the objective (context-invariant) notions of sensitivity and strength of epistemic position. For Lewis, there are the objective factors of truth, belief, reliability, elimination of possibilities, and resemblance, though the latter is better described as only part objective, since only *salient* resemblance is relevant. According to Williams, by contrast, the constraints on truly ascribing knowledge across contexts are many and various, with the result that there is nothing at all to serve as an object of theoretical investigation for the epistemologist.

Part and parcel of repudiating skepticism, then, is repudiating traditional epistemology. Both rely for their livelihood on the assumption Williams calls "epistemological realism," viz. that there are objective relations of epistemic priority waiting to be described.

Paul Moser identifies three projects for epistemology: the *semantic* project of specifying what it means to say that something is justified, the *explanatory* project of identifying informative explanatory conditions that state the conditions for justification, and the *evaluative* project of formulating standards for determining whether a belief is justified. He then considers a dilemma for the evaluative and the explanatory projects. Thus, he writes (here regarding the evaluative project): "Whatever my evaluative epistemic standards are, I must ask: in virtue of what, if anything, are my standards an *adequate* solution to the evaluative

Introduction

project, at least for myself?" In other words, what makes my standards the *right* standards for judging justified beliefs? The dilemma is this: Either I take this question seriously or I don't. If I don't, I'm criticizable as being meta-epistemically naive, for the question is a perfectly legitimate one. If I do take it seriously, I will be seemingly forced into giving circular answers such as "My standards are correct because the verdicts they deliver about justification are *likely* to be true or *probably* true."

Moser claims that no such dilemma blocks success in the semantic project. A definition purporting to capture what one means by "justification" is adequate just in case it does capture what one means. One can then set about to find out what one means. Moser maintains, further, that semantic considerations "hold the key to avoiding the other manifestations of the dilemma of naivete or circularity." What one means by "adequate for evaluating justified beliefs" is what constitutes correctness for solutions to the evaluative project, and similarly for the case of "adequate for explaining justification."

Moser puts these ideas to work in formulating what he calls *semantic foundationalism*, the view according to which answers to questions about correct standards for explaining, evaluating, and arguing for justification properly end in considerations about one's operative notion of justification, i.e., what one means by "justification." Thus, he formulates the following reasoning to be used to argue for one's justification for a statement *P*:

1. By my conceptual commitment, justification consists in conditions *C*.
2. The statement that *P* satisfies *C*.
3. Hence, the statement that *P* is justified (for me).

Relativism enters into Moser's account in the following way. *My* notion of justification may differ from yours, and yours from another's and so on. This raises the practical question of which of these notions one should employ. About this, there is no objectively right *epistemic* answer. It is not as if some, but not all, concepts of justification latch on to epistemic natural kinds or epistemic Platonic forms. Rather, one must weigh the advantages and disadvantages of employing the various concepts in light of one's goals, whatever they may be.

This last theme is also sounded by Stephen Stich in his contribution to this section. Stich's principal object is to examine the implications for

analytic epistemology of *the problem of cognitive diversity*. The problem is this: In light of the many distinct cognitive processes which we might have employed and which others do employ, why should we remain with *ours*? Or, more generally: How can we determine which of the many available cognitive processes are *good* ones? Analytic epistemology traditionally sought to give us the tools to answer these normative questions by armchair analysis of our concept of justification. Stich argues that it cannot do this.

Before Stich attempts to indict analytic epistemology in general, however, he attacks a method endorsed by many analytic epistemologists for identifying which inferential processes are good ones. The method is the method of reflective equilibrium, whereby we attempt to bringing our intuitive judgments about the applicability of a concept in particular cases in line with our general beliefs about the conditions under which the concept applies. Stich asks about the connection between *justification* and *the conditions under which "justified belief" applies as determined by the method of reflective equilibrium*: are they one and the same? The answer must be yes, if the results of reflective equilibrium are to be of use to us in our normative project. But empirical results have shown us that this cannot be right. Ordinary subjects, for example, standardly affirm the gambler's fallacy in reflective equilibrium, and yet inferences conforming to this fallacy are not justified.

Stich then considers the analytic epistemologist's likely response: "When we make judgments applying our concept of justification, *something* guides in making these judgments. There *has* to be some uniform set of conditions that we somehow rely upon. Admittedly, reflective equilibrium isn't quite the method we use to find out what these conditions are. But the conditions are there, and we rely on them, and so there must be some armchair method available for finding them out." Stich replies, once again, by citing empirical results that purportedly show that we do not, even implicitly, consult necessary and sufficient conditions in applying our general concepts, but rather employ in our judgments prototypes and exemplars.

Stich's second argument attempts to show that even if, contrary to what Stich's first argument would lead us to expect, analytic epistemology can determine the necessary and sufficient conditions under which our concept of justification applies, analytic epistemology can provide no

help in solving the problem of cognitive diversity. For analytic epistemology can at best tell us only what *our* concept of justification is. It doesn't tell us that the cognitive processes declared justified by *our* concept are better than those declared justified by *other* concepts of justification. To solve

the problem of cognitive diversity, we must consider how well adopting various concepts of justification would serve our general goals. *That* is not a matter for armchair inquiry, but for empirical investigation.

Further Reading

Moser, Paul K., *Philosophy After Objectivity* (New York: Oxford University Press, 1994).

Putnam, Hilary, *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981).

Rorty, Richard, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979).

—, *Consequences of Pragmatism* (Minneapolis: University of Minnesota Press, 1982).

—, *Contingency, Irony, and Solidarity* (New York: Cambridge University Press, 1989).

—, "Putnam and the Relativist Menace," *The Journal of Philosophy* 90 (1990), pp. 443–61.

—, "Antirepresentationalism, Ethnocentrism, and Liberalism," in Rorty, *Objectivity, Relativism, and Truth: Philosophical Papers, vol. 1* (New York: Cambridge University Press, 1991), pp. 1–17.

Siegel, Harvey, *Relativism Refuted: A Critique of Contemporary Epistemological Relativism* (Dordrecht: Reidel, 1987).

Sosa, Ernest, "Serious Philosophy and Freedom of Spirit," *The Journal of Philosophy* 84 (1987), pp. 707–26.

Stich, Stephen, *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation* (Cambridge, MA: Massachusetts Institute of Technology Press, 1990).

Williams, Michael, *Groundless Belief* (New Haven, CT: Yale University Press, 1977).

—, *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism* (Cambridge, MA: Blackwell, 1991).

Epistemological Realism

Michael Williams

Generality and Epistemic Priority

Although a defender of the naturalness of sceptical doubts must hold that foundationalism is a by-product of scepticism, not a presupposition, so far we have seen nothing to suggest that the case for scepticism can be understood apart from the doctrine of the priority of experiential knowledge over knowledge of the world. This result would not be decisive if this essential doctrine could itself be derived from the truistic elements in the sceptic's arguments. But we have seen nothing to suggest this either. On the contrary, everything points the other way.

This leaves one option: to see how the truistic elements in the sceptic's arguments take on sceptical significance, we must look to the distinctive character of the traditional epistemological project. The sceptic (or traditional epistemologist) must argue that, in the context of a distinctively philosophical investigation of our knowledge of the world, the crucial ideas about epistemic priority are *forced* on us by our ordinary understanding of knowledge or justification. If he can do so, he will have rebutted the charge that he simply takes them for granted.

In trying to explain how what might otherwise seem to be truisms take on a surprising significance, it is natural to look first to the traditional epistemologist's aim of assessing the *totality* of our knowledge of the world. Because he wants to explain how we are able to know anything at all

about the external world, his plan is to assess all such knowledge, all at once. But surely, the argument now goes, if we are to understand how it is possible for us to know *anything at all* about external reality, we must trace that knowledge to knowledge we should still have even if we knew nothing about the world. No explanation of how we come to have knowledge of the world that depended on our already having some would show the required generality: it would not be an explanation of how we have *any* such knowledge. But this is as good as to say that, once we accept the legitimacy of the epistemologist's question – and we have seen no reason to suppose that it is unintelligible – we must also accept the priority of experiential knowledge, since experiential knowledge is what remains when knowledge of the world is set aside.

This is Stroud's view, which explains why he thinks that the diagnosis of scepticism that traces it to foundationalism gets things upside down. According to Stroud:

What we seek in the philosophical theory of knowledge is an account that is completely general in several respects. We want to understand how any knowledge at all is possible – how anything we currently accept amounts to knowledge. Or less ambitiously, we want to understand with complete generality how we come to know anything in a certain specified domain.¹

It is the distinctively philosophical goal of understanding certain kinds of knowledge with "complete generality" that leads to attempts to ground knowledge of a given kind on some "epistemologically prior" kind of knowledge, and the reason is

Originally published in M. Williams, *Unnatural Doubts* (Oxford: Blackwell Publishers, 1991), pp. 83–93, 101–19, 121–4, 129–39.

that no other strategy will yield the right kind of generality. Unfortunately, the lesson of scepticism seems to be that such attempts are bound to fail, so that there is no hope of understanding human knowledge in general.

We can characterize the unusual generality of the traditional epistemological undertaking by saying that the traditional epistemologist imposes a *totality condition* on a properly philosophical understanding of our knowledge of the world. Acceptance of this condition, I believe, is what lies behind the feeling that arguments concerning conceptual points are unfair to the sceptic. Purely conceptual points – the neutrality of experience or the “non-dreaming” implication of ordinary perceptual knowledge – have no intrinsic epistemological significance. Moreover, since such sceptical significance as they possess depends entirely on a tacit commitment to the priority of experiential knowledge over knowledge of the world, they themselves give no grounds for accepting any such general relation of epistemological priority. But perhaps they do not have to. Perhaps the very nature of epistemological investigation forces us to recognize that relation; and once it is recognized, the sceptic’s truistic conceptual points are all he needs to reach his conclusion.

For example, one might argue that the (truistic) claim that my knowing (perceptually) that *P* implies my knowing that I am not dreaming that *P* is not equivalent to the claim the sceptic must assimilate it to: that my knowing that *P* requires my being able to rule out the possibility that I am dreaming that *P* independently of my knowledge that *P* (or indeed anything like it). But the suggestion now is that the totality condition, rather than the non-dreaming condition alone, is what imposes the crucial restriction. So, in the context of the traditional attempt to understand our knowledge of the world, an otherwise innocuous claim gives the sceptic what he needs.

Acceptance of the totality condition on a properly philosophical understanding of our knowledge of the world is also the deep source of the epistemologist’s dilemma, for the dilemma springs from a fatal interaction of the totality condition with the objectivity requirement. This is the requirement that the knowledge we want to explain is knowledge of an objective world, a world that is the way it is independently of how it appears to us to be or what we are inclined to believe about it. Now, as we have seen, the totality condition requires us to try to trace our knowledge of the world to some-

thing more fundamental, which can only be experiential data. But, as a sceptical argument along Ayer’s lines reveals, it is impossible to explain how such data could ever function as evidence. They cannot be linked empirically with any facts about the world for, in accepting such linkage, we would be crediting ourselves with knowledge of the world, in violation of the totality condition. On the other hand, conceptual connections between experiential data and worldly fact seem to be ruled out by the familiar thought-experiments that the sceptic appeals to to establish the neutrality and autonomy of experience. And if, in a desperate attempt to avoid scepticism, we insist on such connections, we make the way the world is depend on how it appears to us, in violation of the objectivity requirement. Accordingly, in the context of the attempt to assess the totality of our knowledge of the world, it seems impossible either to respect or violate the objectivity requirement: whatever we do looks like succumbing to the sceptic.

Nevertheless, although the epistemologist’s dilemma arises from the interaction of the totality condition and the objectivity requirement, I take the totality condition to be fundamental. Many philosophers would disagree, for they see the objectivity requirement, with its commitment to a “realistic” view of truth, as the deep source of sceptical problems. But it is not clear, to me at least, that the objectivity requirement, any more than its relative the neutrality of experience, has any particular sceptical potential outside the context of an assessment of worldly knowledge governed by the totality condition.

I say that the totality condition is fundamental. More strictly, however, what is fundamental is the attempt to conduct an *assessment* of our knowledge of the world in the light of that condition. If the priority of experiential knowledge over knowledge of the world is implicit in the traditional epistemological project, this is not solely on account of that project’s unusual generality. Also crucial is the kind of understanding it suggests we seek. As Quine has argued, if all we want is some kind of causal or developmental account of the emergence of our knowledge of the objective world, there is nothing viciously circular in our appealing to what we now know about the world in an explanation of how we came to be in our current position.² And where there is no threat of circularity, there is no pressure to accede to a general doctrine of epistemic priority.

As Quine is of course well aware, traditional epistemology is under pressure to accept such a doctrine because it seeks a different kind of understanding. Its aim is to explain how it is that our beliefs about the world amount to knowledge. Thus when Stroud says that what we want from a theory of knowledge is an account of how our knowledge of the world emerges out of something that is not our knowledge of the world,³ he does not mean that we want an explanation of how our current way of looking at things developed out of some some previous way: i.e. out of knowledge (or what our ancestors thought of as knowledge) that is not *ours*. This is a task for historians and anthropologists. Nor is he thinking of an account of how our knowledge emerges out of something that is not our *knowledge*. Quine's idea of a naturalized epistemology is a gesture in this direction, for it is supposed to issue in a causal explanation of how our interactions with the environment lead us to form certain beliefs; and if there is a worthwhile project here, it is presumably one for psychologists and neurophysiologists. What is missing from both these projects is the idea of an assessment. Each could as well, in fact more properly, be offered as an account of the emergence of our *beliefs*. But only a legitimating account of the basis or emergence of our beliefs will give an account of our *knowledge*. The sort of theory Stroud has in mind is therefore one that traces our knowledge of the world to something that is *ours*, and that is *knowledge*, but not knowledge of *the world*. What could this be except experiential knowledge? Even Quine is forced to something like this position when he tries to connect his "naturalized" epistemology with traditional sceptical problems.

It seems, then, that something very like foundationalism falls out of a methodological constraint on a properly philosophical examination of knowledge of the world. So we have, apparently, found what we were looking for: a defence of the claim that foundationalism is a by-product of scepticism, not a presupposition. When this possibility was first mooted, I suggested that it would have to turn out that scepticism and foundationalism have a common root. We have now located that common root in the attempt to gain a certain kind of understanding of our knowledge of the world. In effect, we have glossed Hume's thought that we set foot on the road to scepticism as soon as we ask distinctively philosophical questions about knowledge. True, this will not yield a defence of the

naturalness of sceptical doubts unless, as Hume thought, that form of questioning is itself fully natural. However, even on this point, the sceptic has strengthened his position. It is hard to see how there could be anything *unintelligible* in what seems only to be an attempt to understand knowledge in an unusually general way, so the prospects for a convincing therapeutic diagnosis of scepticism seem bleak. But it is not obvious offhand that the prospects for a satisfactory theoretical diagnosis are any brighter, for how can mere generality entail extensive theoretical commitments?

This is not all. Suppose that we agree that the traditional epistemological project leads inevitably to the conclusion either that we have no knowledge or that, if we do, we will never understand how we do; and suppose we insist that, since this is its outcome, it *must* involve some distortion of our epistemological position: can we say that identifying this distortion will let us see how knowledge is possible after all? Stroud suggests not. We should not think that:

if we did come to see how and why the epistemological enterprise is not fully valid, or perhaps not even fully coherent, we would then possess a satisfactory explanation of how human knowledge in general is possible. We would have seen, at best, that we cannot have any such thing. And that too, I believe, would leave us dissatisfied.⁴

This is a powerful objection to any theoretical diagnosis of scepticism. Attempts to answer the sceptic directly run into the epistemologist's dilemma. But, if Stroud is right, attempts at diagnostic responses meet a similar fate. Suppose we find that we cannot hope to ground our knowledge of the world in the way that traditional epistemology has invited us to, because of some defect in the ideas about justification involved in the notion of even trying: we would still not have explained to ourselves how it is that we ever come to know anything about the world. Unless we show that the sceptic's question is actually unintelligible, it will remain dissatisfyingly unanswered. So this is our new dilemma: if the traditional epistemological project is coherent, it is doomed to fail; and if it isn't, we are still left in a position hard to distinguish from scepticism. It may be scepticism at second order, but it is scepticism for all that. We may *have* knowledge of the world, but we will never be able to explain to ourselves how we do.

We may know things about the world, but we will never know that we know them.

Knowledge as an Object of Theory

In asking whether there is such a thing as knowledge of the world, I am not asking the very same question the sceptic asks but one that I think cuts deeper. I am asking how we have to think about “knowledge of the world” for that phrase to pick out a proper object of theory. So if it sounds too strange even to hint that there might not be anything for the theory of knowledge to be a theory of, my question can be rephrased. What matters is whether “our knowledge of the world” picks out the kind of thing that might be expected to be susceptible of uniform theoretical analysis, so that failure to yield to such analysis would reveal a serious gap in our understanding.

To raise these questions is to begin to examine a move that gets made before epistemological arguments, and particularly sceptical arguments, even get started. This is the introduction of the objects of epistemological inquiry. We shall be trying to isolate views that, for the most part, even the most determined anti-sceptics share with their adversaries. Philosophers who respond to scepticism do not doubt that there is something to defend against the sceptic’s attacks. If they are dubious about our prospects for giving a direct refutation of scepticism, they call for a diagnosis of the sceptic’s questions which will reveal them, first impressions to the contrary, as less than fully coherent. Even Stroud, who thinks our most pressing need as epistemologists is to understand how traditional epistemological inquiry misrepresents our epistemic position, if it does, seems not to doubt the existence of its objects. For the idea that there is something called “our epistemic position” is just another aspect of the idea that there is such a thing as “human knowledge” or “our view of reality.” But is there? Or are there fewer things in heaven and earth than are dreamt of in our epistemology?

Now, it is tempting to use “human knowledge” and “our knowledge of the external world” as though it were obvious that such phrases pick out reasonably definite objects of study. But it isn’t obvious, or shouldn’t be. We can talk of “our knowledge of the world,” but do we have any reason to suppose that there is a genuine totality here and not just a loose aggregate of more or less

unrelated cases? My sense is that the totality condition is far more problematic than it first seems.

Consider, for example, Nagel’s characterization of the aim of epistemology as “to form a conception of reality which includes ourselves and our view of reality among its objects.”⁵ This offhand allusion to “our view of reality” takes a lot for granted. To suppose that there is such a thing as “our view of reality,” which might then be the “object” of a single theoretical enterprise, is to assume that human knowledge constitutes some kind of surveyable whole, an idea that is not, on the surface, very promising. There are no clear criteria for individuating beliefs and, even if there were, it is far from clear that there would be any systematic way of enumerating all the things we believe. Phrases like “our system of beliefs” and “our view of reality” are so vague that we cannot be confident they refer to anything.

Nothing changes if we pull back to narrower categories such as knowledge of the external world. When it comes to such “specified domains,” whether there is anything to understand will depend on how the domains are specified. To try to understand all knowledge in the standard epistemic domains is to suppose that the beliefs in those domains hang together in some important way. But how? “Knowledge of the external world” covers not only all the natural sciences and all of history, it covers all everyday, unsystematic factual claims belonging to no particular investigative discipline. Since, even within a single subject, theories, problems and methods tend to proliferate with the progress of inquiry, so that even the most systematic disciplines tend to become less rather than more unified, it is doubtful whether we can take a synoptic view of physics, never mind everything we believe about the external world. It is not obvious that it makes sense even to try.

Recall Stroud’s claim that in the philosophical study of human knowledge we want “to understand how any knowledge at all is possible – how anything we currently accept amounts to knowledge.” He finds that engaging in this project “feels like the pursuit of a perfectly comprehensible intellectual goal.”⁶ Perhaps it does once we have grown familiar with theoretical ideas that we shall be examining shortly. But we must try to recover some naivete here. Then I think we see that, when we first encounter the challenge to show how any knowledge of the world is possible, we cannot tell whether we have been given a perfectly

comprehensible goal or not. In fact, the obvious difficulty in commanding a synoptic view of our worldly beliefs suggests that we haven't. We cannot, therefore, just *see* whether the epistemological challenge makes sense. What we can do, however, is to ask how we might make sense of it.

I think that we can find a somewhat oblique recognition of this problem even in Descartes. Descartes admits that getting to a general doubt by questioning his beliefs one at a time would not be easy: perhaps the examination would never be completed. Hume too dismisses a piecemeal approach as a "tedious lingering method."⁷ But these grudging concessions are misleading: for they imply that the main obstacles to going over our beliefs *seriatim* are time and energy, whereas the question is certainly not one of convenience. If we are to make sense of the project of explaining how anything we believe about the world amounts to knowledge, we need a way of reducing our beliefs to order. We have to bring them under principles or show them as resting on commitments that we *can* survey. We must reveal some kind of *theoretical integrity* in the class of beliefs we want to assess.⁸ If we can do this, human knowledge is a possible object of theoretical investigation. But not otherwise.

The very nature of the traditional project demands that the principles in question be all-pervasive. For example, if we are to assess the totality of our beliefs about the world, there must be principles that inform all putative knowledge of the world *as such*. But what could they be? I take it to be obvious that, in one way, our beliefs do not show any kind of theoretical integrity. They do not, that is, add up to an ideally unified theory of everything. There is no way now, and none in prospect, of integrating all the sciences, much less all of anyone's everyday factual beliefs, into a single coherent system: for example, a finitely axiomatized theory with specified rules of inference. In this way, Nagel's phrase "our view of reality" borders on the absurd. We have not got a "view of reality" but indefinitely many. The idea, taken for granted by coherence theorists of justification, that we have a "system" of beliefs ought to be suspect.

"Our beliefs," then, do not amount to a single, integrated "view of reality." They are not *topically integrated*. But this need not be fatal to the project of understanding human knowledge in general. For even if our beliefs are not topically integrated, they might be *epistemologically* integrated. This is

to say: they might be subject, in so far as they are meant to be justified or to amount to knowledge, to the same fundamental, epistemological constraints. This is what is usually suggested, or rather assumed. Thus Descartes ties his pre-critical beliefs together, thereby constituting their totality as an object of theoretical inquiry, by tracing them all to "the senses." No matter how topically heterogeneous, and no matter how unsystematic, his beliefs have this much in common: all owe their place to the authority of the senses. If this authority can be called in question, each loses its title to the rank of knowledge.

We have seen that this talk of "the senses" is poised between a causal truism and a contentious epistemological doctrine. Now we see more clearly why the epistemological doctrine is and must be what is intended. Only by tracing our beliefs about the world to a common "source," which is to say a common evidential ground, can we make "beliefs about the world" the name of a coherent kind. In the absence of topical integration, we must look to epistemological considerations for the theoretical integrity we require.

Hume may have seen, though perhaps dimly, that an epistemologically based form of theoretical integrity is a precondition for a properly general, hence "philosophical," understanding of human knowledge. He compares assessing particular beliefs and particular sciences one at a time to a strategy of "taking now and then a castle or village on the frontier"; and he contrasts this "tedious" method with marching up to "the capital or center of these sciences, to human nature itself." In explaining the principles of human nature, he tells us, "we in effect propose a compleat system of the sciences." But the completeness envisaged does not involve topical integration. It derives rather from the fact that all sciences, whatever their subject matter, "lie under the cognizance of men, and are judged of by their powers and faculties." Their subjection to the same underlying epistemological constraints, rooted in our "powers and faculties" is thus what makes possible a sweeping evaluation of "all the sciences."⁹

Hume sees the fact that all sciences lie "under the cognizance of men" as showing that all are "in some measure dependent on the science of MAN." But it seems clear that the science of man is not, or ought not to be, dependent on the other sciences. (Hume is apologetic about his occasional excursions into natural philosophy.) This asymmetry belongs to the logic of Hume's project, indeed to

the logic of the traditional epistemological enterprise. Since he is attempting, with a view to its reform, a wholesale assessment of our knowledge of both the physical and the moral world, he cannot take any of that knowledge for granted. This means that it must be possible to investigate our "powers and faculties," the epistemological aspect of the human condition, without relying on any worldly knowledge. Our epistemological self-knowledge must be both autonomous and fundamental. Thus the project of assessing the totality of our knowledge of the world does more than presuppose that experiential knowledge is in some very deep way prior to knowledge of the world. It also assigns a definite privilege to knowledge of such epistemological facts. These features of the traditional project point to very extensive theoretical commitments.

The fact that the traditional epistemological enterprise is committed to the autonomy of epistemology sheds further light on the significance of externalism in the theory of knowledge. By suggesting that our capacity for knowledge depends on our situation in the world, and not just on our own "internal" capacities, externalism challenges the idea of our "epistemic position" as an autonomous object of theory. If our epistemic position is not something that can be investigated without knowing something about how we are placed in the world, there can be no question of our assessing the totality of our knowledge of the world on the basis of insights into our epistemic position. Perhaps we do not even have a fixed epistemic position. And if we find that we do not, it is doubtful whether we will be able to retain a clear conception of "our knowledge of the world" as an appropriate object of theory.

Unlike Hume, Descartes aspires to topical as well as epistemological integration: hence his metaphor of the tree of knowledge whose roots are metaphysics, trunk physics, and branches medicine, mechanics, and morals, a figure that contrasts interestingly with Hume's citadel of reason. But even for Descartes, topical integration is something to be achieved rather than assumed. His initial survey of his beliefs takes for granted only their epistemological integrity. As is familiar, he makes the point in terms of the metaphor of foundations: undermine the foundations and the whole edifice crumbles. The metaphor is a very natural one for, as we have seen, there is a clear sense in which epistemology, understood as the attempt to comprehend how any knowledge is

possible, is intrinsically foundational. To see human knowledge as an object of theory, we *must* attribute to it some kind of systematic basis. This may involve inference from some class of fundamental evidence-conferring beliefs, as traditional foundationalists maintain; or it may involve governance by certain "global" criteria of explanatory integration, as coherence theorists think. But *something* must regulate our knowledge of the world: something that we can identify and examine independently of any such knowledge. We should therefore not be too eager to *oppose* the account of scepticism that traces it to the generality of the epistemological enterprise to that which traces it to foundationalism. (Nor, for that matter, should we be too eager to oppose foundationalism to the coherence theory.) If we give up the idea of pervasive, underlying epistemological constraints; if we start to see the plurality of constraints that inform the various special disciplines, never mind ordinary, unsystematic factual discourse, as genuinely irreducible; if we become suspicious of the idea that "our powers and faculties" can be evaluated independently of everything having to do with the world and our place in it: then we lose our grip on the idea of "human knowledge" as an object of theory. The clear contrast between castles on the frontier and the fortress at the centre dissolves. Perhaps there is no capital, each province, as Wittgenstein said of mathematics, having to take care of itself. The quest for an understanding of human knowledge as such, no longer feels like "the pursuit of a perfectly comprehensible intellectual goal."

The same is true of more modest aims, such as understanding how our beliefs about the external world amount to knowledge. As a way of classifying beliefs, "beliefs about the external world" is only quasi-topical, bringing together beliefs belonging to any and every subject, or no well-defined subject at all. They are united only by their supposed common epistemological status. The essential contrast to "beliefs about the external world" is "experiential beliefs" and the basis for the contrast is the general epistemic priority of beliefs falling under the latter heading over those falling under the former. "External" means "without the mind"; and it is taken for granted that we have a firmer grasp of what is "in" the mind than of what is outside it.

There is no doubt that this epistemological distinction is readily mastered: readily enough for arguments based on it to strike us as "immediately

gripping.” But a teachable distinction does not guarantee theoretical integrity in the kinds of things distinguished. There are various ways of failing. I discuss two examples in this section and one in the next.

My first example illustrates a relatively mild form of failure. In his natural history of heat, Bacon gives a long list of examples of heating. It includes examples of heating by radiation, friction, exothermic reactions, and by “hot” spices that “burn” the tongue.¹⁰ Everything he mentions is ordinarily said to involve “heat,” so we cannot deny that his list reflects ordinary usage. But what we have here is a clear case in which a nominal kind, comprising all the things commonly called “hot,” has no automatic right to be considered a natural kind. It is no objection to the kinetic theory that it doesn’t cover the tremendous “heat” produced in my mouth by a chicken vindaloo, never mind the heat often generated by philosophical arguments. We don’t complain that, since the theory doesn’t apply to hot curries or heated arguments, it fails to explain heat in a satisfactorily general way.

Given that we want to know whether there is any such thing as, say, “our knowledge of the world,” this kind of failure may seem too weak to be of interest. Failure to take in hot curries and heated arguments does not tempt us to say that there is no such thing as heat. But we could say that there is no such thing as *nominal* heat, the nominal kind being *merely* nominal. We can tie together some of the examples of heat and, having done so, treat them as the only genuine examples, discarding the others as resembling the genuine examples only superficially, hence as not really, but only metaphorically, hot. This is, indeed, what Bacon himself goes on to do when he argues that heat is a form of motion. Anyway, it is clear that there need be no theory of all the things commonly called “hot”: a hot curry is hot even when it has gone cold. Nor need the lack of such a theory be cause for intellectual dissatisfaction. It is just another example of an ordinary principle of classification failing to cut nature at the joints. By the same token there does not *have* to be a theory of all the things normally called “examples of knowledge.” And if there isn’t, it has to be shown that this reveals a lack. It may be that there is no such thing as knowledge (or knowledge of the external world, etc.) in just the way that there is no such thing as Bacon’s nominal heat.

All this notwithstanding, I agree that the example of heat doesn’t get me very far. All that happens in this case is that a nominal kind fails to coincide exactly with a theoretically coherent kind. So I move to my second example: the supposed division of sentences into analytic, or true by virtue of meaning, and synthetic, or true by virtue of fact. Quine is famously sceptical about this distinction because he is dubious about the atomistic conception of meaning that he takes to lie behind it.¹¹ Quine’s view of meaning is holistic – the meaning of a given sentence depends on its role in a wider theory – and this holistic conception of meaning suggests that there is no privileged way of distinguishing a theory’s “meaning postulates” from its empirical assumptions, any more than there is a way of determining which out of alternative complete axiom sets is the right one.

Against Quine, Grice and Strawson argue that the analytic/synthetic distinction must be genuine and significant because it is teachable in such a way as to enable the student to apply it to new cases.¹² The reply, well known by now, is that all kinds of dubious distinctions have proved to be teachable in this way, for even terms belonging to a false theory can admit of consensual application on the part of those who accept it. If the fact that, at one time, everyone could agree on who was the village witch does not mean that there really were witches, the fact that appropriately trained students can pick out examples of analytic sentences does nothing to show that any sentences are genuinely analytic. But the point I want to make does not require agreement on this particular example. Whether or not we agree with Quine on the question of analyticity, the fact remains that distinctions can be teachable and projectible while failing to correspond to any theoretically coherent division of objects. When a classification rests on an implied background theory, there is no immediate inference from the existence of an easily mastered kind-term to the theoretical integrity of its associated kind.

The application to our current problem is obvious. In accordance with my project of theoretical diagnosis, I have been arguing that the kinds of knowledge investigated by the traditional epistemologist are theoretical kinds. So, just as the ability of believers in the analytic/synthetic distinction to agree on what to count as paradigm instances of analytic sentences does not mean that there are analytic sentences, the fact that we can agree on what to count as examples of knowledge

of the external world does not mean that there is knowledge of the external world. The underlying principle of classification, whatever it is, might be bogus. As a result, we cannot simply help ourselves to classifications of this kind on the grounds that nothing else promises the right kind of generality. That such principles of classification pick out coherent objects of theoretical investigation needs to be shown.

In the case of heat, to sort out the genuine from the spurious examples we rely on a physical theory which identifies some underlying property, or structure of more elementary components, common to hot things. Explaining theoretically significant kinds this way is typical of scientific realism. For the scientific realist, deep structural features of the elementary components of things determine the boundaries of natural, as opposed to merely nominal or conventional, kinds. This suggests an analogy. Since, if human knowledge is to constitute a genuine kind of thing – and the same goes for knowledge of the external world, knowledge of other minds, and so on – there must be underlying epistemological structures or principles, the traditional epistemologist is committed to *epistemological realism*. This is not realism as a position within epistemology – the thesis that we have knowledge of an objective, mind-independent reality – but something quite different: realism about the objects of epistemological inquiry.

The epistemological realist thinks of knowledge in very much the way the scientific realist thinks of heat: beneath the surface diversity there is structural unity. Not everything we call knowledge need be knowledge properly so called. But there is a way of bringing together the genuine cases into a coherent theoretical kind. By so doing – and only by so doing – we make such things as “knowledge of the external world” the objects of a distinctive form of theoretical investigation. We make it possible to investigate knowledge, or knowledge of the world, *as such*.

I expect that at first it seemed bizarre to question the existence of the objects of epistemological inquiry. Who can deny that we evaluate claims and beliefs epistemologically, sometimes deciding that they express or amount to knowledge, sometimes not? And who can deny that these claims or beliefs concern such things as objects in our surroundings, other peoples’ thoughts and experiences, events in the past, and so on? No one. So it is easy to assume that, if our claims ever warrant positive assessment, there must be knowledge of

the external world, knowledge of other minds, knowledge of the past, and so on. Even more obviously, there must be knowledge. But I hope the examples just considered make plausible the thought that there doesn’t *have* to be. All we know for sure is that we have various practices of assessment, perhaps sharing certain formal features. It doesn’t follow from this that the various items given a positive rating add up to anything like a natural kind. So it does not follow that they add up to a surveyable whole, to a genuine totality rather than a more or less loose aggregate. Accordingly, it does not follow that a failure to understand knowledge of the world with proper generality points automatically to an intellectual lack. To sum up, though I readily admit that we have teachable distinctions here, all this ensures is that there will be things that we can agree on as *examples* of, say, knowledge of the external world. It does not guarantee any theoretical integrity of the kind to which the examples are assigned. This is the sense in which there need be no such thing as knowledge of the world.

At this point, someone is likely to object that there is no immediate inference from the lack of a certain type of theoretical integrity in a given kind to its spuriousness. Still less is there an inference to the non-existence of things of that kind. Take the sort of loose, functional classification of things that is common in everyday life, such as the division of dining room furniture into table and chairs. We do not expect to be able to formulate a physical theory of what makes an object a chair. But we are not tempted to conclude that chairs do not exist.¹³

This objection assumes that “knowledge of the external world” is like “chair” rather than like “witch” (or “analytic”). But is it? The distinctive feature of terms like “witch” is that they are *essentially theoretical*. Essentially theoretical distinctions are distinctions that we see no point in continuing to make, or even no way of drawing, once the theory behind them has been rejected. If Quine is right, “analytic/synthetic” is like this, for he holds that giving up a certain conception of meaning involves losing all sense of how to make a fixed, objective division between a theory’s meaning postulates and its empirical assumptions. Essentially theoretical classifications must therefore be distinguished from classifications that have been theoretically rationalized but which retain independent utility. Distinctions like this are apt to survive the rejection of theories with which they have become associated. Our first

example, heat, is a case in point. Rejecting the caloric theory of heat, or the phlogiston theory of combustion, did not tempt us to conclude that there are no hot things or that nothing burns. Some philosophers would take this view of “analytic,” for they think that there is a robust and useful pre-theoretical notion of synonymy that survives Quinean scepticism about meanings. If they are right, the analytic/synthetic distinction is not essentially theoretical. But where a classification is essentially theoretical, we are happy to say that there are no things of that kind, if we once become convinced that the background theory is false. Thus there are no witches (or, if Quine is right, analytic sentences).

Though I do not claim that the concept of an essentially theoretical classification is knife-edged, I do want to say that “knowledge of the external world” is quite clearly essentially theoretical. There is no commonsense, pre-theoretical practice that this way of classifying beliefs rationalizes: its sole function is to make possible a certain form of theoretical inquiry, the assessment of knowledge of the world as such. As we have seen, this classification cuts across all familiar subject-matter divisions and, in addition, presupposes the autonomy of epistemology. Even the sense of “external” is unfamiliar from a commonsense standpoint. “External” does not mean “in one’s surroundings,” for even one’s own body, with its “internal organs,” is an “external” object. It was a radical innovation on Descartes’s part to externalize his own body.¹⁴ As I have already remarked, “external” in “external world” means “without the mind.” And since being within the mind depends on being given to consciousness, the essential contrast to “knowledge of the external world” is “experiential knowledge”: the classification is epistemological through and through.

But what if the proper analogy for “knowledge of the external world” were not “witch” but “heat?” I do not believe that it is because I do not see that there is any pre-theoretical utility to the concept, or any theory-independent way of drawing even approximately the right boundaries round it. But this is not all. In bringing to centre-stage the issue of epistemological realism, I am not questioning particular theories of the structure of empirical knowledge, as we might question particular theories of heat, but the very idea that knowledge has any fixed, context-independent structure. The analogy is therefore not with cases where one structural theory replaces another but with those where we

abandon any idea of coming up with a theory of that kind. If there are no witches, we may debate witch-crazes and witchcraft beliefs, but not whether sympathetic magic is superior to contagious.

Suppose, however, that I am wrong about all this. Suppose, that is, that “knowledge of the external world” is like “chair”: then what? So far as I can see, nothing to the purpose. In connection with such loose, functional classifications, we do not expect theoretical understanding, which is why such classifications survive the recognition that no such understanding will be forthcoming. We do not feel that there is an irremediable intellectual lack because there will never be a science of chairs. But that is exactly what we are supposed to feel in the absence of a suitably anti-sceptical theory of knowledge of the external world. This shows that, even by the traditional epistemologist’s own standards, “knowledge of the external world” cannot be like “chair.” It must pick out something in which theoretical integrity is to be expected, and this means that the existence of the objects of traditional epistemological inquiry is far less assured than that of furniture.

Explanation or Deflation?

Let me suggest one further case for comparison. It has to do with deflationary views of truth. Philosophers who take a deflationary approach want no more from a theory of truth than a description of the logical behaviour of “true” and some account of why it is useful to have such a device in our language. Quine is a good example of such a philosopher. According to Quine, if we consider a sentence like “‘Snow is white’ is true if and only if snow is white” we see that: “To ascribe truth to the sentence is to ascribe whiteness to snow. ... Ascription of truth just cancels the quotation marks. Truth is disquotation.”¹⁵

Applied to a given sentence, the truth-predicate is dispensable. It comes into its own, however, with respect to sentences that are not given, as when we say that all the consequences of a given theory are true. But even here, to say that certain sentences are true is just to say that the world is as *they* say it is. As Quine remarks, “one who puzzles over the adjective ‘true’ should puzzle rather over the sentences to which he ascribes it. ‘True’ is transparent.”¹⁶

Though I am very sympathetic to this view, my interest here is less in its correctness than its

character. This view of truth is striking on account of what it does *not* say. Compared with traditional theories of truth, it says nothing about what makes all true sentences true. On the contrary, a deflationist will hold that his remarks on the behaviour and utility of the truth-predicate say just about everything there is to say about truth. To approach truth in a deflationary spirit is emphatically not to think of “true” as denoting a theoretically significant property, explicating which will illuminate what is involved in any sentence’s being true. What is involved in a given sentence’s being true is exhaustively captured by the sentence itself. On a deflationary view, then, true sentences constitute a merely nominal kind. We could even say that, for a deflationist, though there are endlessly many truths, there is no such thing as truth.

The traditional theorist sees things quite differently. In his eyes, “truth” is the name of an important property shared by all true sentences, a property that can be expected to repay theoretical analysis. This property may be correspondence to fact, incorporability in some ideally coherent system of judgments, or goodness in the way of belief, depending on whether he favours a correspondence, coherence, or pragmatic theory. But whatever his theoretical preference, he will hold that, since true sentences constitute not just a nominal but a theoretical kind, no theory of truth is satisfactory which does not explain what makes true sentences true. We set our sights too low if we aim only to capture the use of a word or explain the point of a concept: there is more to understanding truth than appreciating the utility of the truth-predicate.¹⁷

We see, then, that traditional and deflationary theories are not theories of exactly the same kind. As Stephen Leeds puts it, the traditional theories are genuinely theories of truth whereas deflationary theories are theories of the concept of truth (or, we could say, accounts of the use of “true”).¹⁸ Leeds’s illuminating distinction is readily applied to epistemological theories. We can distinguish theories of knowledge from theories of the concept of knowledge. I think that the debate sparked by Gettier’s demonstration that the standard “justified true belief” analysis fails to state a sufficient condition for knowledge is best seen as concerning the concept of knowledge. The kind of extra constraint on justification that seems to be required – for example that an inference cannot yield knowledge if it involves a false lemma essentially – is rather formal, nothing being said about what beliefs can serve as justifying evidence for what.

This is why it is possible to discuss issues raised by the Gettier problem without ever getting entangled in sceptical problems. Theories that say nothing about whether examples of justified beliefs about objective states of affairs reveal any essential similarities, beyond highly formal ones of the “no false lemmas” variety, are neutral with respect to whether we should think of our knowledge of the world as an appropriate object of theory. By contrast, traditional foundational and coherence theories, which are much more closely involved with scepticism, put forward general, substantive constraints on justification and so make room for a project of assessing our knowledge of the world as a whole. They are theories of knowledge and not just theories of the concept of knowledge.¹⁹

Of course, there is no obstacle in principle to supplementing one’s views about the concept of knowledge with views about knowledge itself.²⁰ But one could also advance such views in a deflationary spirit. One philosopher who has done so, I believe, is Austin. Wittgenstein may be another.

The availability of deflationary accounts of a notion like truth changes the whole problem-situation. Naively, we might be inclined to suppose that just as in physics we study the nature of heat, so in philosophy we study the nature of truth. But once plausible deflationary views are on the table, the analogy between truth and things like heat can no longer be treated as unproblematic, for the question raised by such views is precisely whether there is any need to think of truth as having a “nature.” We can conclude, *mutatis mutandis*, that if we have a plausible account of the concept of knowledge, it is a further step to insist on an account of knowledge as well. A deflationary account of “know” may show how the word is embedded in a teachable and useful linguistic practice, without supposing that “being known to be true” denotes a property that groups propositions into a theoretically significant kind. We can have an account of the use and utility of “know” without supposing that there is such a thing as human knowledge.

What makes this suggestion particularly pointed is that appearances certainly do not favour the view that a phrase like “knowledge of the world” picks out a theoretically coherent kind. For one thing, justification, like explanation, seems interest-relative, hence context-sensitive. This is in part what Austin is driving at in insisting that demands for justification are raised and responded to against a background of specifically relevant error

possibilities. What is relevant will depend on both the content of the claim in question and the context in which the claim is entered. If all evidence is relevant evidence, then, abstracting from such contextual details, there will be no fact of the matter as to what sort of evidence could or should be brought to bear on a given proposition.

If context-sensitivity goes all the way down, there is no reason to think that the mere fact that a proposition is “about the external world” establishes that it needs, or is even susceptible of, any particular kind of evidential support. No proposition, considered in abstraction, will have an epistemic status it can call its own. To suppose that it must be precisely to fall in with what I call “epistemological realism.” To treat “our knowledge of the world” as designating a genuine totality, thus as a possible object of wholesale assessment, is to suppose that there are invariant epistemological constraints underlying the shifting standards of everyday justification, which it is the function of philosophical reflection to bring to light. Exposing this epistemological deep structure will be what allows us to determine, in some general way, whether we are entitled to claim knowledge of the world. But if this is so, foundationalist presuppositions are buried very deeply in the Cartesian project. They do not just fall out of the totality condition’s exclusion of any appeal to knowledge of the world in the course of our attempt to gain a reflective understanding of that knowledge. They turn out to be involved in the very idea of there being something to assess.

These are my suspicions in outline. Now we must look at some details.

Foundationalism

My main concern is the relation between scepticism and foundationalism. So having distinguished between theories of knowledge and theories of the concept of knowledge, I must say what kind of a theory I take foundationalism to be.

One way to understand foundationalism is to see it as a doctrine about the formal character of justifying inferences. Formal foundationalism, as we may call it, is the view that justification depends on the availability of terminating beliefs or judgments, beliefs or judgments which amount to knowledge, or which are at least in some way reasonably held to, without needing support from further empirical beliefs. Formal foundationalism

is sometimes thought to contrast with “coherentist” theories of knowledge or justification. According to theories of this type, a given belief becomes justified through incorporation in some suitably “coherent” system of beliefs or “total view.” Empirical inference is thus a matter of moving from one total view to another. The terminating judgments, which the foundationalist sees as fixed points constraining the possibilities of inferential justification, are unnecessary. Some philosophers see the commitment to beliefs that function as fixed points as *the* essential feature of foundationalism, hence the complaint, prominent in a recent systematic defence of the coherence theory, that the key error in foundationalism is its “linear” conception of inference.²¹

I have my doubts about the contrast between foundationalism and the coherence theory, but they can wait. The point I want to make here is that anyone who traces scepticism about our knowledge of the external world to the foundationalist doctrine of epistemic priority must have more than formal foundationalism in mind. We can call this stronger doctrine “substantive” foundationalism. The distinction between formal and substantive foundationalism turns on the account given of terminating beliefs or judgments. Substantive foundationalism involves more than the formal doctrine that inference depends on letting certain beliefs function as fixed points: it adds a distinctive account of the kind of beliefs capable of performing that function. Since I think that a genuinely foundationalist view of knowledge and justification must be substantive, whenever I refer to foundationalism *simpliciter* I shall have substantive foundationalism in mind.²²

Substantive foundationalism is a theory of knowledge, whereas formal foundationalism is only (a contribution to) a theory of the concept of knowledge. One way to see this is to recall that Wittgenstein’s view of knowledge, which concedes that all justification takes place against a background of judgments affirmed without special testing, can be seen as *formally* foundationalist. But this point about our ordinary practices of justification, while it might offer a way into the fully general problem of the regress of justification, gives no basis for supposing that there is a particular sceptical problem about our knowledge of objective reality. The transition to that problem depends on the tacit assumption that the fixed points recognized by commonsense justifications fall into some fairly obvious kind, so that once they

have been questioned there must be some other, more primitive kind of judgment that we are forced to look to for their support. The thought that the functional role recognized by formal foundationalism corresponds to some kind of broad topical division of our beliefs is what I take to be the essential characteristic of substantive, as opposed to merely formal, foundationalism.

This is the way, then, in which there is more to what I am calling (and what has generally been called) "foundationalism" than the purely structural doctrine of formal foundationalism. What is missing from formal foundationalism is any hint as to the kinds of beliefs that function as fixed points or as to what qualifies a belief to play that role. But we have not yet got quite to the heart of why formal foundationalism is too weak a doctrine to capture all that is essential to a foundationalist conception of knowledge and justification. The key point is this: that not only does formal foundationalism give no account of what sorts of beliefs are epistemologically prior to what, and why, it does not even imply that any such account needs to be given. If foundationalism is a *purely* formal or structural doctrine, we have no reason to think that a given belief has *any* particular or permanent epistemological status. Perhaps the same belief can be a fixed point at one time, or in one particular context of inquiry or justification, but a candidate for justification at another time or in another context. Nothing in formal foundationalism excludes this.

By contrast, substantive foundationalism presupposes epistemological realism. I first introduced the idea of epistemological realism by way of analogy with scientific realism. We can now get a clearer sense of the appropriateness of the analogy. A micro-structural theory of a physical phenomenon is not purely structural. It will identify both certain structures and the types of entities fitted to occupy appropriate places in them. (Think of models of the atom.) Similarly with the foundationalist: he both attributes to justifying inferences a certain structural character *and* identifies the types of beliefs fitted to play the various structurally defined roles: basic, inferential, etc. Thus for the (substantive) foundationalist beliefs have an *intrinsic epistemological status* that accounts for their ability to play one or other of the formal roles the theory allows. Beliefs of one kind can be treated as epistemologically prior to beliefs of some other kind because they *are* epistemologically prior; some beliefs play the role of basic beliefs

because they *are* basic; others receive inferential justification because they *require* it; and all because of the kinds of beliefs they are. According to foundationalists, our beliefs arrange themselves into broad, theoretically coherent classes according to certain *natural* relations of epistemological priority. Beliefs to which no beliefs are epistemologically prior are epistemologically basic. Their credibility is naturally intrinsic, as that of all other beliefs is naturally inferential. This is a much more peculiar doctrine than is generally recognized.²³

On the foundationalist view, a belief's intrinsic epistemological status derives from the content of the proposition believed. The foundationalist's maxim is "Content determines status." Not, however, the details of content: what matter are certain rather abstract features, for example that a belief is about "external objects" or "experience." Thus it comes naturally to foundationalists to talk of basic propositions or basic statements, as well as of basic beliefs. Propositions recording the data of experience are held to be, by their very nature, epistemologically prior to propositions about external objects, which is why they are apt for the expression of basic beliefs. In light of this, we can characterize foundationalism as the view that our beliefs, simply in virtue of certain elements in their contents, stand in *natural epistemological relations* and thus fall into *natural epistemological kinds*. The broad, fundamental epistemological classes into which all propositions, hence derivatively all beliefs, naturally fall constitute an epistemic hierarchy which determines what, in the last analysis, can be called on to justify what. This means that, for a foundationalist, every belief has an inalienable epistemic character which it carries with it wherever it goes and which determines where its justification must finally be sought. The obvious illustration is the thought that any belief whatever about "external objects" must in the end derive its credibility from the evidence of "the senses," knowledge of how things appear.

I call the foundationalist's supposed relations of epistemological priority "natural" to emphasize the fact that they are supposed to exist in virtue of the nature of certain kinds of beliefs and not to depend on the changing and contingent contexts in which beliefs become embedded. For the foundationalist, in virtue of his epistemological realism, there is a level of analysis at which epistemic status is not, as Quine once said of one important epistemic feature, conventionality, "a passing trait."

Beliefs are more like the members of a highly class-conscious society in which a person, no matter what he does, always carries the stigma or cachet of his origins. The quest for epistemic respectability is thus never entirely *une carrière ouverte aux talents*. A given belief, though useful in all sorts of ways, generally and quite properly (in appropriate contexts) taken for granted, and beyond any specific reproach, can never be allowed quite to forget that it *presupposes the existence of the external world* and is therefore, by that fact alone, subject to some kind of residual doubt, unless it can trace its lineage to more respectable data.

The foundationalist conception of fundamental epistemological relations, cutting across ordinary subject divisions and operating independently of all contextual constraints, receives an early articulation in Descartes's notion of "the order of reasons." Descartes writes, "I do not follow the order of topics but the order of arguments.... [In] orderly reasoning from easier matters to more difficult matters I make what deductions I can, first on one topic, then on another."²⁴ However, it is far from obvious that there is such an order of reasons, operating independently of the division of topics. It is not at all clear that some matters are intrinsically – that is to say independently of all circumstances and all collateral knowledge – "easier" than others. The way that justification and inquiry proceed in common life, or for the matter theoretical science, is far from evidently favourable to the foundationalist conception of epistemological relations. In both science and ordinary life, constraints on justification are many and various. Not merely that, they shift with context in ways that are probably impossible to reduce to rule. In part, they will have to do with the specific content of whatever claim is at issue. But they will also be decisively influenced by the subject of inquiry to which the claim in question belongs (history, physics, ornithology, etc.). We can call these *topical* or, where some definite subject or distinctive form of inquiry is involved, *disciplinary* constraints. Not entertaining radical doubts about the age of the Earth or the reliability of documentary evidence is a precondition of doing history *at all*. There are many things that, as historians, we might be dubious about, but not these.

Disciplinary constraints fix ranges of admissible questions. But what is and is not appropriate in the way of justification may also be strongly influenced by what specific objection has been entered to a given claim or belief. So to disciplinary we must

add *dialectical* constraints: constraints reflecting the current state of a particular argument or problem-situation. In this respect justification is closely akin to explanation, which is also context-sensitive because question-relative.

I shall have more to say about disciplinary constraints and about the relation between justification and explanation. But for now let me note that, in ordinary examples of requiring and producing justifications, the epistemological status of a given claim can also depend on the particular situation in which the claim is entered, so that justification is also subject to a variety of *situational* constraints. Here I have in mind the worldly and not just the dialectical situation. Consider yet again Wittgenstein's remark that "My having two hands is, in normal circumstances, as certain as anything I could produce in evidence for it."²⁵ Entered in the right setting, a claim to have two hands might function like a foundationalist's basic statement, providing a stopping place for requests for evidence or justification: hence the element of formal foundationalism in Wittgenstein's view. But in other circumstances *the very same claim* might be contestable and so might stand in need of evidential support. The content of what is claimed does not guarantee a claim some particular epistemic standing. Not merely is status often dependent on the details of content, it is *never* determined by content alone. As Wittgenstein notes:

If a blind man were to ask me "Have you got two hands?" I should not make sure by looking. If I were to have any doubt of it, then I don't know why I should trust my eyes. For why shouldn't I test my *eyes* by looking to find out whether I see my two hands? *What* is to be tested by *what*? (Who decides *what* stands fast?)²⁶

The point is that, in the absence of a detailed specification of a particular context of inquiry, the sort of specification that would fix the relevant contextual constraints on justification, the question "What is to be tested by what" has no answer. Questions about justification are essentially context-bound. This is something a foundationalist will deny. He must of course make allowances for the way that what tests what can shift with context. But – and this is the crucial point, he cannot allow that such contextual determination goes all the way down. At the fundamental level, what is to be

tested by what is objectively fixed, which is why there is no question of anybody's *deciding* the matter. The answer is determined by the epistemological facts themselves: by fundamental, objective relations of epistemological priority. This is not exactly an "intuitive" view.

Continuing with the example of my knowing (in normal circumstances) that I have two hands, recall also that there is no obvious way to generalize from an example like this. In normal circumstances, the proposition that I have two hands is as certain as anything we could cite as evidence for it. But there is no obvious, non-trivial way of saying what other propositions are, in normal circumstances, as certain as anything we could cite as evidence for them. Normally, I am as certain as I could be of anything that my name is Michael Williams: but beyond this, what does the proposition that my name is Michael Williams have in common with the proposition that I have two hands? What feature of their content explains their belonging to the same epistemic kind? As far as I can see, there isn't one. So even if someone said that the claim to have two hands did have a kind of intrinsic status – that of being certain in normal circumstances – we would still not be able to treat the example as *paradigmatic* of propositions belonging to a definite epistemic kind, for which we could articulate some alternative, non-trivial criterion of membership.²⁷ Again, the foundationalist sees things quite differently. For him, highly abstract divisions of propositions according to content (propositions about external objects versus experiential propositions, propositions about the past versus propositions about the present, etc.) have to coincide with fixed differences in epistemological status. But what we should learn from the example under discussion is that no such coincidence can be simply assumed. To cite again another of Wittgenstein's reminders, "a proposition saying that here is a physical object may have the same logical status as one saying that here is a red patch."²⁸ Without natural epistemological kinds, the foundationalist's permanent underlying structure of epistemological relations goes by the board.

We see from this that the antidote to foundationalism, indeed to epistemological realism generally, is a *contextualist* view of justification.²⁹ To adopt contextualism, however, is not just to hold that the epistemic status of a given proposition is liable to shift with situational, disciplinary, and other contextually variable factors: it is to hold

that, independently of all such influences, a proposition has no epistemic status whatsoever. There is *no fact of the matter* as to what kind of justification it either admits of or requires. Thus stated, contextualism implies a kind of externalism, for though appropriate contextual constraints will have to be met, if a given claim is to express knowledge, they will not always have to be known, or even believed, to be met.³⁰ But when we realize that the point of contextualism is to oppose the sceptic's or traditional epistemologist's epistemological realism, the externalist element in contextualism ought to be more palatable. The problem with externalism was that it seemed to deprive us of the possibility of answering a perfectly intelligible question: how do we come to know anything whatsoever about the external world? What we now see is that this question is not at all intuitive but reflects theoretical presuppositions that are not easy to defend. Contextualism, with its implied externalism, is not offered as a question-begging direct answer to an undeniably compelling request for understanding, but as a challenge to justify the presumption that there is something to understand.

Methodological Necessity

We have already seen that, to flesh out the idea of "human knowledge" as a possible object of theoretical investigation, we have to suppose that there are pervasive epistemological constraints or relations. That is to say, at least some constraints on what propositions demand evidential support and on what propositions can provide it must be *context-invariant*. If we do not always insist on respecting these constraints in a fully rigorous way, this need not mean that they do not apply. To admit that certain constraints are often waived is different from, indeed incompatible with, claiming that they are inapplicable.

This is a very substantial commitment and it is not clear why we should accept it. An examination of ordinary practices of justification strongly suggests that constraints, governing what sorts of evidence can properly be brought to bear on a disputed claim, what needs to be defended, and what can safely be taken for granted, though subject to other kinds of contextual determination as well, are at least *topic-relative*, which is to say determined in part by the subject under discussion.

We might criticize Hume's offhand suggestion that only carelessness and inattention save us from a permanent, debilitating awareness of the truth of scepticism, hence from lapsing into a state of chronic, paralysing doubt. In particular contexts, disciplines etc., exempting certain propositions from doubt is what determines the *direction* of inquiry. As Wittgenstein remarks: "It may be . . . that all enquiry on our part is set so as to exempt certain propositions from doubt, if they are ever formulated. They lie apart from the route travelled by enquiry."³¹

If some of these propositions cease to lie apart from the route travelled by inquiry, then inquiry travels by a different route. Or perhaps no clear route remains for it to travel by. This is obviously the case with investigations in particular scientific or scholarly disciplines. Disciplinary constraints have a great deal to do with the kinds of questions that can and cannot legitimately be raised without radically affecting the direction of inquiry. Thus, introducing sceptical doubts about whether the Earth really existed a hundred years (or five minutes) ago does not lead to a more careful way of doing history: it changes the subject, from history to epistemology. So when Wittgenstein asks: "am I to say that the experiment which perhaps I make to test the truth of a proposition presupposes the truth of the proposition that the apparatus I believe I see is really there?"³² he is clearly inviting the answer "No." And the reason for answering "No" is that the possibility mentioned, while relevant to certain general, epistemological problems, is completely beside the point in the context of a specific experiment in chemistry or physics. To bring it up is not to introduce greater rigour into the investigation in hand but to shift attention to another kind of investigation entirely.

"[T]hat something stands fast for me," Wittgenstein remarks, "is not grounded in my stupidity or credulity."³³ We now see that this is so, at least in part, because it is grounded in my *interests*. It is not that I think that no proposition that stands fast could ever be questioned, though in certain cases I should be likely to feel, as Wittgenstein says, "intellectually very distant" from someone inclined to raise questions. It is just that some doubts are logically excluded by forms of investigation that I find significant, important, or perhaps just interesting. This has nothing to do with dogmatism, credulity or carelessness. Wittgenstein sums up the key points in the following well-known passages:

The questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn.

That is to say, it belongs to the logic of our scientific investigations that certain things are indeed not doubted.

But it isn't that the situation is like this: We just can't investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinges must stay put.³⁴

Of course, if I do not want the door to turn I can nail it shut; or I might want it to open the other way, in which case I will move the hinges. But if I want the door to turn this way, it is not just more *convenient*, if a little slapdash, to place the hinges where they are: there is nowhere else to put them.

By fixing a range of admissible questions, we determine a form of inquiry. But this means that a form of inquiry is determined by more than purely formal constraints. As Wittgenstein puts it: "The question doesn't arise at all.' Its answer would characterise a *method*. But there is no sharp boundary between methodological propositions and propositions within a method."³⁵ For a subject like history, there is more to method than abstract procedural rules. This is because the exclusion of certain questions (about the existence of the Earth, the complete and total unreliability of documentary evidence, etc.) amounts to the acceptance of substantial factual commitments. These commitments, which must be accepted, if what we understand by historical inquiry is to be conducted at all, have the status, relative to that form of inquiry, of *methodological necessities*.

I have introduced the idea of a proposition's being exempted from doubt as a matter of methodological necessity in connection with the disciplinary constraints that determine the general directions of highly organized forms of inquiry. But it is evident that something similar goes on in more informal, everyday settings. Asking some questions logically precludes asking others: all sorts of everyday certainties have to stand fast if we are to get on with life. Again, however, I want to emphasize that our situation is misread both by the Humean naturalist and by the sceptic. The naturalist sees our everyday inability to entertain radical doubts as showing that nature has simply determined us to believe certain things, however groundless they seem to us in our more reflective

moments. By contrast, I want to claim that exemption from doubt – epistemic privilege – is a matter of methodology, not psychology. In a specific context, certain exemptions will be *logically* required by the direction of inquiry. We are therefore determined by Nature to hold certain things fast only in so far as we are naturally inclined to interest ourselves in matters requiring us to exempt them from doubt.

This is far from the only point that we must emphasize. It is also crucial to note that, if epistemic status is determined by the direction of inquiry, the reason why, in a given inquiry, certain propositions have to stand fast has to be separated from the reason why that inquiry results in knowledge, if it does. Here we recur, from a slightly different angle, to the externalist element in contextualism. In particular contexts of inquiry, certain propositions stand fast as a matter of methodological necessity. But inquiries informed by them will yield knowledge only if those propositions are true, which they need not always be.

The general moral here is that questions about a proposition's epistemic status must always be separated from questions about its truth. If epistemic status is fixed by the direction of inquiry, epistemic status is context-sensitive. Truth however is not. A proposition is either true or not. But, according to the contextualist view I favour, we cannot say, in a similarly unqualified way, that a proposition is either open to doubt or not. Sometimes it will be and sometimes it won't. Generally speaking, a proposition is neither true because it stands fast nor stands fast because it is true.

We can also see why it was so important at the outset to distinguish between formal and substantive foundationalism. If foundationalism is equated with a certain view of the formal structure of justification – i.e. with the view that inferential justification always requires beliefs that function as "fixed points" – a contextualist view of justification can be seen as (formally) foundationalist. But it certainly need not be substantively foundationalist. There are no limits as to what might or might not, in an appropriate context, be fixed.

In an earlier chapter, I tried to show that arguments for radical scepticism presuppose the priority of experiential knowledge over knowledge of the world. This enabled me to conclude that attempts to establish the intrinsic epistemological priority of experiential knowledge on the basis of the greater intrinsic dubitability of objective knowledge are question-begging. The only reason

for thinking that such knowledge is intrinsically more dubitable is provided by the existence of sceptical arguments which, when unpacked, turn out to take the doctrine of the priority of experiential knowledge for granted.

This result did not allow us to conclude straight away that scepticism rests on a gratuitous epistemological assumption. What it did suggest, however, is that the source of the doctrine of the priority of experiential knowledge is not evidence from our ordinary justificational practices but rather the distinctively philosophical project of trying to understand how it is possible for us to know anything whatsoever about the external world. The totality condition that the sceptic (or the traditional philosopher) imposes on a philosophical understanding of our knowledge of the world is what forces us to see that knowledge as somehow derivative from experience. No other way of seeing it permits an assessment, hence a legitimating explanation, at the proper level of generality.

We are now in a position to see why this argument does not prove what it needs to prove. All it shows is that the doctrine of the priority of experiential knowledge over knowledge of the world is a *methodological necessity of the traditional epistemological project*. But since the sceptic himself is irrevocably committed to distinguishing between methodological necessity and truth, it does not show, nor by his own standards can the sceptic take it to show, that that doctrine is true.

The result is that the inference from the essential generality of the traditional epistemological project fails to establish the kind of relations of epistemological priority needed to threaten us with scepticism. To yield sceptical results, these relations must reflect more than *mere* methodological necessities: they must correspond to fully objective epistemological asymmetries. It is not enough to point out that if we are to attempt an assessment of our knowledge of the world as a whole we must *take* experiential knowledge to be epistemologically prior to the knowledge we want to assess. Success or failure in the enterprise will have the significance the sceptic and the traditional epistemologist mean it to have only if experiential knowledge really is, as a matter of objective epistemological fact, more basic than knowledge of the world. If it isn't, or more generally if no epistemological relations are in the sense I have indicated fully objective, no attempt to ground knowledge of some allegedly problematic kind on some

appropriately prior kind of knowledge will amount to an attempt at assessment. Should the attempt fail, or even inevitably fail, the sceptic will be left with a harmless logico-conceptual point but with no way of advancing to his pessimistic epistemological conclusion.

I remarked that the argument from the totality condition to the absolute priority of experiential knowledge over knowledge of the world rests on two assumptions: that there is something to assess, and that charting its relation to experience amounts to assessing it. I have concentrated on the first, but by so doing have shown what to say about the second. As a pure methodological proposal, there is nothing wrong with setting propositions about the world against experiential propositions, for the purposes of exploring possible relations between them. Like Goodman, we could think of phenomenalism as an interesting constructive project. We could ask, "To what extent can a phenomenalist reconstruction of the world be carried through?" without thinking that we were even addressing any questions of epistemic legitimacy.³⁶ Think of the way we can model arithmetic in set theory: though this is an interesting piece of mathematics, we need ancillary epistemological assumptions to think of it as relevant to an "assessment" of arithmetic. But this is not the spirit in which the sceptic thinks of the relation between experiential knowledge and knowledge of the world. He needs a fully objective epistemological asymmetry, and this is what no argument from methodological necessity will ever yield.

Some philosophers, Carnap for example, hold that the sceptic fails to undermine ordinary knowledge of the world because his statements, as he intends them to be taken, mean nothing at all. As a statement "internal" to our everyday linguistic framework, "There are material objects" is a trivial consequence of any statement about the world. But as an "external" statement about that framework, an attempted statement, though made in the very same words, will lack "cognitive significance." However, the sceptic might be equally unsuccessful if his statements, as they must be understood in the unusual context of philosophical reflection, mean something different from what they ordinarily mean. Thus Thompson Clarke suggests that the very general commonsense propositions with which Moore confronts the sceptic can be taken two ways, the "plain" way and the "philosophical" way. For example:

Suppose a physiologist lecturing on mental abnormalities observes: *Each of us who is normal knows that he is now awake, not dreaming or hallucinating, that there is a real public world outside his mind which he is now perceiving, that in this world there are three-dimensional animate and inanimate bodies of many shapes and sizes.* ... In contrast, individuals suffering from certain mental abnormalities each believes that what we know to be the real public world is his imaginative creation.³⁷

The italicized, plain propositions are "verbal twins" of propositions typically attacked and defended in discussions of philosophical scepticism. But in plain contexts, nobody doubts that they are true, even though plain common sense recognizes the very phenomena – dreaming, hallucinating, and so on – that the sceptic appeals to in his attempt to show that we can never know that we are in touch with "a real public world." Whether there is a clash between philosophy and common sense will depend, therefore, on the relation between philosophical and plain knowing.

Here Clarke is more subtle than Carnap, for he recognizes that the sceptic has an account of the relation between philosophy and common sense which both preserves the relevance of philosophical discoveries to ordinary plain knowing and makes it hard to think that sceptical claims are less than fully meaningful.³⁸ Ordinary, plain knowing is hemmed by practical considerations. By contrast, to philosophize is "to step outside the nonsemantical practice" and, meaning simply what one's words mean, ask whether we really know what we (plainly) take ourselves to know. Compared with our philosophizing, ordinary thinking is "restricted." All the sceptic has to do is to get us to look beyond the restrictions. This is easy enough since there is a standing invitation to look "beyond the plain" in our conception of knowledge as knowledge of an objective world. We want to know what there is: not just relative to this and that particular restriction, imposed by this or that practical purpose or limitation, but *absolutely*.

Still, the final distance between Clarke and Carnap is not as great as their initial divergence might suggest. Clarke too holds that, in the end, both "philosophical common sense" and its sceptical denial "are a spurious fiction if our conceptual-human constitution is not standard." Amongst other things, a conceptual-human constitution of

the standard type requires that “Each concept or the conceptual scheme must be divorceable intact from our practices, from whatever constitutes the essential character of the plain” and that we, as concept users, are “purely ascertaining observers who, usually by means of our senses, ascertain, when possible, whether items fulfill the conditions legislated by concepts.”³⁹ But the sceptic himself shows that our conceptual-human constitution cannot be of the standard type. Our plain knowledge that we are not dreaming right now – the sort of knowledge expressed by the physiologist – cannot be undermined by the plain possibility that we might, in fact, be asleep. But it would be if our conceptual-human constitution were of the standard type. For on this point the sceptic is right: there are no marks or features that conclusively distinguish waking experience from dreaming. So the fact of plain knowing, combined with the sceptic’s point about dreaming or hallucinating, shows that our conceptual-human constitution is not of the standard type. This insight is part of the legacy of scepticism.

In representing the sceptic as helping bring about his own undoing, Clarke prefigures the strategy followed by Wright. Wright, we may recall, argues that the sceptic does indeed show that his target-propositions – for example, that there is a real, public world – are beyond justification. They are beyond justification because the sole evidence we can bring to bear on them only functions as evidence if they are already known to be true. Thus sensory experience only counts in favour of any proposition about the public world on assumptions that already commit us to that world’s existence. But the lesson to learn from this is that the propositions the sceptic represents as groundless, factual assumptions, are not really factual at all. If a proposition’s factuality requires some account of the cognitive powers that would be required for knowing that proposition to be true, and if the sceptic shows that, in the case of some propositions, no such account can be given, scepticism is self-undermining. This argument shares with Clarke’s more than just structural similarities.

None of these arguments appeals to me. I do not want to distinguish between internal and external questions or between plain and philosophical meanings of statements. Nor do I wish to claim that, for deep philosophical reasons, apparently factual statements are really not factual at all. The reason is that I think that all these reactions

to scepticism reveal the deep and pervasive influence of epistemological realism. I suggested earlier that one of the epistemological realist’s central commitments is to the doctrine that content determines status. Now I claim that the attempt to insulate common sense from sceptical undermining by finding a different meaning, or no factual meaning at all, in the apparently commonsensical propositions the sceptic examines is driven by that same doctrine. If a statement is certain in one context but not in another, the argument assumes, this can only be because a change in context induces a change in meaning. So if, plainly speaking, we do know that we are awake at the moment, whereas, philosophically speaking, we don’t, our plain and philosophical propositions can only be “verbal twins.” But if, as I have argued, epistemological status is never determined by content alone, there is no such easy inference from a difference in status to a difference in content. We can explain the context-boundedness of sceptical doubts without getting entangled in this baroque apparatus of plain and philosophical meanings. As we shall see in a moment, this is all to the good.

Once again, I must emphasize that my argument on these matters will not be complete until I have examined the sceptic’s own favoured account of the nature of philosophical reflection. Even so, however, I think it is fair to conclude that we are well on the way to accomplishing the primary goal of theoretical diagnosis, which is to get the sceptic to share the burden of theory. But there is a nagging question that is likely to surface again at this point. If we are left with one theory of knowledge confronting another, and we will never be able to determine conclusively which is correct, doesn’t the sceptic win ties and so triumph at second order?

If we abandon epistemological realism, there is a clear sense in which we no longer see such things as “knowledge of the world” as appropriate objects of theory. At most, we will have a theory of the concept of knowledge. We will not have a theory of knowledge as well. *A fortiori*, we will not be left confronting the sceptic’s theory with a theory of our own.

Perhaps this will look like a purely verbal manoeuvre, for we shall certainly be left with epistemological views, whether or not we want to think of them as a theory of knowledge. But the point isn’t just verbal. For what we have seen is that the sceptic’s theoretical commitments are in fact far more extensive than those of his contextualist

opponent. Contextualism simply takes seriously and at face-value what seem to be evident facts of ordinary epistemic practices: that relevant evidence varies with context, that content alone never determines epistemological status, and so on. The theoretical resources required to explain these appearances away belong entirely to the sceptic. So it might be reasonable to object that the sceptic wins ties, if the outcome of my theoretical diagnosis were a tie. And if I had followed philosophers like Carnap, Clarke, or Wright and rested my diagnosis on difficult and controversial views about meaning, perhaps it would have been. But as things stand it isn't.

This is not all. It seems to me entirely reasonable to hold that extra theoretical commitments demand extra arguments. But where will the sceptic find them? Not in evidence from everyday practice, which fits in as well or better with contextualism. Presumably, then, in some kind of general, theoretical considerations. Here, however, we run into the fallaciousness of the argument from methodological necessity: by the sceptic's own standards, there is no inference from the fact that we must *take* experiential knowledge to be generally prior to knowledge of the world, if we are to make room for a project of assessing our knowledge of the world as a whole, to its really being so. But if the argument from methodological necessity does not show that the sceptic's principles are true, what would? It is hard to say: for

although the argument from methodological necessity is fallacious, it is not as if there are other ways of arguing for the priority of experiential knowledge. On the contrary, as we have seen repeatedly, attempts to argue for it directly beg the question. So the doctrine has to be true but unarguable.

I think that the sceptic's difficulties are compounded when we turn from this relatively particular doctrine to epistemological realism in general. It is not easy to imagine what a convincing argument for epistemological realism would even look like, or what evidence it could appeal to. This is where the clash between scepticism and our ordinary attitudes really does work to the sceptic's disadvantage. It does so because our ordinary practices of justification not only tolerate but invite a contextualist construction: and contextualism is the antidote to epistemological realism.

True, a contextual view of knowledge and justification will seem unsatisfactory to a philosopher who continues to feel the lack of an understanding of human knowledge in general. But if my argument to this point is correct, he will feel this lack only if he is already predisposed to epistemological realism. Once more, we are starting to run round a very small circle of ideas. The sceptic's foundationalism, together with the epistemological realism it embodies, is a brute metaphysical commitment. The theoretical diagnostician could hardly ask for more.

Notes

- 1 Barry Stroud, "Understanding Human Knowledge in General," in Marjorie Clay and Keith Lehrer (eds), *Knowledge and Skepticism* (Boulder, Co: Westview, 1989), p. 32.
- 2 W. V. Quine, "Epistemology Naturalized," this vol., ch. 23. On Quine's problematic attitude towards traditional epistemology, see Barry Stroud, *The Significance of Philosophical Skepticism* (Oxford: Oxford University Press, 1984), ch. VI.
- 3 Barry Stroud, "Skepticism and the Possibility of Knowledge," *Journal of Philosophy* 81, pp. 545–51. p. 551.
- 4 "Understanding Human Knowledge," p. 49.
- 5 Thomas Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1986), p. 68.
- 6 "Understanding Human Knowledge," p. 32.
- 7 René Descartes, "Meditations on First Philosophy: First Meditation," in *The Philosophical Works of*

- Descartes*, vol. 1, trans. Elizabeth Haldane and G. R. T. Ross (Cambridge: Cambridge University Press, 1972), p. 145; David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd edn rev. P. H. Nidditch (Oxford: Oxford University Press, 1978), p. xx (hereafter, *THN*).
- 8 I am grateful to Simon Blackburn for this useful phrase.
- 9 These quotations and the next, *THN*, pp. xv–xvi.
- 10 *Novum Organum*, bk II in J. Spedding, R. Ellis, and D. Heath (eds), *The Works of Francis Bacon* (London: Longman, 1857–8), vol. IV. My example is slightly unfair to Bacon in that he did not intend to give instances of everything that would, in common parlance, be said to involve "heating." His aim was to collect instances which "agree in the same nature, though in substance the most unlike": *Novum Organum*, II, aphorism xi.

- 11 W. V. Quine, "Two Dogmas of Empiricism," in W. V. Quine (ed.), *From a Logical Point of View* (New York: Harper, 1963).
- 12 H. P. Grice and P. F. Strawson, "In Defence of a Dogma," *Philosophical Review* (1956).
- 13 I must thank Alvin Goldman for pressing me on this point.
- 14 Myles Burnyeat ("Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed," *Philosophical Review* 90, pp. 3–40) thinks that Descartes's externalization of his own body is the key move in his invention of the problem of our knowledge of the external world. I see it as a consequence of his epistemological realism.
- 15 W. V. Quine, *Pursuit of Truth* (Cambridge MA: Harvard University Press, 1990), p. 80.
- 16 *Ibid.*, p. 82.
- 17 However, it seems to me that the more a purely disquotational account of "true" can be shown to capture whatever we want out of the truth-predicate, the less reason there is for thinking that there must be some "truth-making" property that all true sentences share: the invitation to apply Occam's Razor ought to be, irresistible.
- 18 Stephen Leeds, "Theories of Reference and Truth," *Erkenntnis* (1978).
- 19 Attempts to discuss the Gettier problem and traditional sceptical questions in the same breath often seem rather contrived. The epistemological analogue of Leeds's distinction explains why.
- 20 Keith Lehrer notes that the "definitional or formal" in his theory of knowledge, which "constitutes an analysis or explication of the concept of knowledge," "leaves open substantive issues." See Lehrer, "Knowledge Reconsidered" in Clay and Lehrer (eds), *Knowledge and Skepticism*, quotation p. 132.
- 21 Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985), pp. 89ff. Subsequent citations given by *Structure* and page numbers.
- 22 The terminology of formal versus substantive foundationalism is also employed by Ernest Sosa: see "The Raft and the Pyramid," this vol., ch. 14. However, I am uncertain whether my usage is the same as Sosa's. According to Sosa, "A type of *formal foundationalism* with respect to a normative or evaluative property ϕ is the view that the conditions (actual and possible) within which ϕ would apply can be specified in general, perhaps recursively. *Substantive foundationalism* is only a particular way of doing so" (p. 278, italics in original). From my point of view, everything depends on what is allowed to count as a "general" specification.
- 23 Though I have grave doubts about the notion of intrinsic credibility, having written about all this elsewhere (*Groundless Belief* (Oxford: Blackwell), chs 2, 3, and 5) I will not repeat myself. My interest here is in the foundationalist's conception of epistemological priority, which I see as his deepest theoretical commitment.
- 24 Letter to Mersenne (24 December, 1640), quoted from Anthony Kenny (ed.), *Descartes: Philosophical Letters* (Minneapolis: University of Minnesota Press, 1981), p. 87.
- 25 Ludwig Wittgenstein, *On Certainty* (Oxford: Blackwell, 1969), p. 250; hereafter *OC*.
- 26 *Ibid.*, p. 125.
- 27 Failure to appreciate this point is what vitiates Marie McGinn's intuitive reconstruction of the case for scepticism.
- 28 *OC*, p. 53.
- 29 For a succinct defence of contextualism, see David B. Annis, "A Contextualist Theory of Epistemic Justification," *American Philosophical Quarterly* (1978), reprinted in Moser, *Empirical Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986). Annis sees Pierce, Dewey, and Popper as having been, historically, the key contextualists. He may be right, though I have doubts about how far these philosophers saw into the implications of contextualism. For example, I doubt whether Popper would be as suspicious as he is about justification if he were really a thoroughgoing contextualist.
- 30 Obviously, this kind of externalism is not the same as pure reliabilism, except in so far as the apparent theoretical simplicity of some forms of reliabilism is a sham. Consider, for example, Colin McGinn's suggestion (in "The Concept of Knowledge," in *Midwest Studies in Philosophy IX* (1984)) that knowing that p depends on the availability of a "way of telling" that p . This analysis does not guarantee any theoretical integrity in "ways of telling" and is therefore compatible with a contextual, hence anti-epistemological-realist, conception of knowledge.
- 31 *OC*, p. 88.
- 32 *Ibid.*, p. 163.
- 33 *Ibid.*, p. 235.
- 34 *Ibid.*, pp. 341–3.
- 35 *Ibid.*, p. 318, emphasis in original. Cf. the metaphor of the river bed at pp. 95–8.
- 36 Nelson Goodman, *Ways of Worldmaking* (Indianapolis: Hackett, 1978), ch. 1.
- 37 Thompson Clarke, "The Legacy of Skepticism," *Journal of Philosophy* 69, pp. 754–69; p. 756.
- 38 *Ibid.*, pp. 758ff.
- 39 *Ibid.*, pp. 762, 761.

Justification, Meta-Epistemology, and Meaning

Paul Moser

1 Three Epistemological Projects

The theory of epistemically justified belief, according to the implicit assumptions of many epistemologists, has at least three main projects:

- (a) the *semantic* project of specifying, in informative terms, what it means to say that something (for example, a proposition or a belief) is epistemically justified;
- (b) the *explanatory* project of identifying informative explanatory conditions that state in nonepistemic terms when, or in virtue of what, a belief is justified; and
- (c) the *evaluative* project of formulating standards for evaluating whether a particular belief is justified

(Section 3 illustrates that another main epistemological project involves arguing for justification.) A proper understanding of how projects (a) through (c) relate to one another and of what constitutes the correctness of their solutions will help to solve some major problems in epistemology and in metaphilosophy. Talk of justification in this chapter will, unless otherwise noted, concern *epistemic* justification, the kind of justification appropriate to propositional knowledge.

The semantic project seeks to define the term 'epistemic justification' (or some synonymous term) by informative synonymous terms. This project, stated thus, is largely neutral on the controversial issue of what exactly constitutes syno-

nymy. Some epistemologists construe the semantic project to require strictly defining conditions that are individually necessary and jointly sufficient for a belief's satisfying the schema "Belief *B* is epistemically justified for person *S*." Epistemologists divide, however, over the issue whether specification of a notion of epistemic justification requires an analysis or a strict definition via necessary and sufficient conditions. They also divide over the question whether we can adequately define "epistemic justification" in nonepistemic terms. (On the latter question, see Moser 1989, pp. 38–44; 1990.)

The explanatory project aims to identify informative explanatory conditions for justified belief. These explanatory conditions will ideally be in nonepistemic terms (at least according to some theorists) and will answer certain explanation-seeking questions about justified belief, questions that purportedly go beyond issues about the meaning of "epistemic justification." (We may now use "explanation" loosely, to encompass answers to "understanding-seeking" what-questions, why-questions, and how-questions.) Explanatory conditions might include considerations about the sources of justified belief or about the extent of justified belief. Pertinent questions include: Are there any nonempirical sources of justified belief? If so, what are they, and how do they yield justification? Does justified belief extend to universal propositions? If so, how? (Some connections between explanation and definition will emerge later in this chapter; for now, we can leave the distinction rough.)

The evaluative project seeks criteria, or guidelines, for deciding whether a belief is justified. It

Originally published in P. K. Moser, *Philosophy After Objectivity* (Oxford and New York: Oxford University Press, 1993), pp. 60–8, 71–82, 86–7, 89–90, 95–9.

aims to go beyond meaning and explanation, to a method for finding out what beliefs are justified: for evaluating whether certain beliefs actually have epistemic justification. Even if we have a definition of “justification” and an explanation of the sources and the extent of justification, we may still lack effective standards for deciding what beliefs are actually justified. Our definitional and explanatory considerations about justification might be too general to serve effectively as guidelines for evaluating the presence of justification in particular cases. The evaluative project, according to some epistemologists, is not necessarily just the application of the results of the semantic or the explanatory project.¹

2 An Epistemological Dilemma: Naivete or Circularity?

For any answer one gives to accomplish the evaluative project, we can raise this question: In virtue of what, if anything, is *that* answer adequate, at least for oneself, for discerning justified beliefs? In other words, in virtue of what, if anything, is that answer a *correct* solution to the evaluative project, at least for oneself? Perhaps one’s solution is not adequate, or correct, at all. Even so, our question can take a modal form: In virtue of what can an answer be adequate, or correct? (For now, “adequate” and “correct” are interchangeable.)

Suppose that I formulate and accept a set of standards for evaluating whether a belief is justified. My evaluative epistemic standards will characteristically take this form:

We may evaluate a (candidate) belief as epistemically justified if (and only if) that belief satisfies conditions *C*.

I might be a familiar empiricist whose evaluative standards invoke conditions regarding best available explanation of one’s experiences; or I might be an equally familiar coherentist whose standards appeal to coherence of some sort among one’s antecedent beliefs and candidates for belief.

Whatever my evaluative standards are, I shall face this question: In virtue of what, if anything, are my standards an adequate, or a correct, solution to the evaluative project, at least for myself? More specifically, what, if anything, constitutes the adequacy, or correctness, of my evaluative

standards as principles for discerning justified beliefs, at least for myself?

Either I take the previous question seriously or I do not. If I do not, I shall disregard it as negligible, as not needing an answer. In that case, my acceptance of my standards for evaluating justification will be naive, or superficial, in at least one obvious respect. My acceptance will then be innocent of a cogent reply to a legitimate question about what, if anything, constitutes the adequacy, or correctness, of my epistemic standards. Call this *meta-epistemic naivete*. It is naivete about answering a question concerning the adequacy of one’s epistemic standards. Such naivete allows for objectionable neglect of intelligible, legitimate questions about what, if anything, constitutes the correctness of one’s epistemic standards.

Neglect here would indeed be objectionable, since affirmation that an epistemic standard, *E*, is a genuine solution to the evaluative project presupposes (for its being fully understood) affirmation that *E* is adequate, or correct, as a solution to that project. Even affirmation that *E* is an epistemically justifiable solution involves a notion of adequacy or correctness; it presupposes a claim to justifiable affirmation that *E* is an adequate, or correct, solution. A notion of correctness is presupposed because full understanding of a notion of a justifiable solution requires an answer to the following question in terms of correctness: Justifiable as what? The answer: As (probably) correct. We need, then, an account of what, if anything, constitutes adequacy, or correctness, of epistemic standards.

If I do take the previous question about adequacy seriously, I shall seek an answer to it. I shall then seek an account of what, if anything, constitutes the adequacy of my evaluative standards for discerning justified beliefs. If *E* is the set of my evaluative epistemic standards, I might appeal to a different set of evaluative epistemic standards, *E'*, to explain what constitutes the adequacy of *E*. This option will ultimately fail, however, for two reasons.

First, *E'* will face a direct analogue of the question facing *E*: In virtue of what, if anything, is *E'* adequate for discerning justified beliefs? Legitimate questions about what constitutes adequacy arise even at higher levels, indeed at every level. It does no good to invoke still another set of evaluative epistemic standards here. Such a pattern of reply will lead only to an endless regress of evaluative standards. One problem is that we finite

humans do not have time to articulate an endless regress of standards when explaining adequacy of epistemic standards. The current reply thus fails as an account of our actually explaining adequacy. Another problem is that an endless answer to our question seems not to be a cogent answer. Such an answer seems too open-ended to offer a resolution of our question.²

The second reason for failure is more decisive. Standards for evaluating merely whether beliefs are justified – even beliefs about the justification of evaluative standards – do not themselves explain what constitutes the adequacy, or correctness, of an evaluative standard. Explanation of mere (evaluable) justification is one thing; explanation of correctness, another. Verificationism about correctness seeks to collapse this distinction. For example, Michael Dummett's verificationism about meaning implies that we "must explain truth as attaching to a statement in some such way as that it does so when the statement either has or could have been verified" (1991, p. 318). Dummett's verificationism is doomed to failure here, because it construes verification in terms of an "acknowledged means of establishing a statement *as true*" (1991, p. 317; italics added). The relevant notion of verification thus presupposes a notion of truth; it does not define, or otherwise explain, that notion in terms of verification.

Typical talk of justification, like Dummett's talk of verification, presupposes talk of justification *as (likely to be) true*, or *as (probably) correct*. Such talk of justification thus does not enable us to sidestep questions about what constitutes correctness. We cannot effectively rely on a notion of correctness – even an implicit notion of correctness – to explain what constitutes correctness. Such a vacuous strategy would make our explanations uninformative and thus pointless. Conceptual circularity is, then, no promoter of effective explanation.

Consider another approach to our question about what constitutes adequacy of epistemic standards. Because, by hypothesis, I already accept a set of standards for discerning what is justified and what is not, I might invoke those very standards to explain what constitutes the adequacy of my standards for discerning justified beliefs. The general idea is this: If *E* is my evaluative standard for discerning justified beliefs, then if *E* meets its own requirements, it is an adequate, or correct, solution to the evaluative project, at least for me.

As noted previously, evaluative standards for discerning simply whether beliefs are justified – even beliefs about the justification of evaluative standards – do not themselves explain what constitutes the adequacy, or correctness, of an evaluative standard. What constitutes adequacy, or correctness, of a standard relative to the evaluative project is one thing; what constitutes mere justification of a standard is something else. Typical talk of justification of a standard presupposes a notion of correctness of a standard, at least insofar as full understanding of such talk requires a notion of likely, probable, or evident *correctness*. Answering questions about what constitutes correctness of standards by appeal to justification, confirmation, or verification of standards thus risks conceptual circularity.

Typical notions of justification, confirmation, and verification presuppose a notion of correctness in need of explanation. They presuppose a notion of correctness inasmuch as full understanding of them requires answers to these questions: Justification as what? Confirmation as what? Verification as what? The presupposed answers are all of one kind: *as (probably) correct*. The verificationism of Dummett and others overlooks such conceptual circularity.

Even if my evaluative standards for discerning justified beliefs might fall short of their own requirements, this question arises: What, if anything, constitutes the adequacy of those requirements for discerning justified beliefs in the first place, aside from what standards actually satisfy those requirements? If the only answer comes from an appeal to the very standards in question, an obvious circle threatens. Clearly, a cogent explanation cannot take this form: My standard *E* is adequate because my standard *E* is adequate. We could use that explanatorily useless form to "explain" the adequacy of whatever we like. The needed explanation, moreover, does not come from a claim that *E* meets its own evaluative standard for discerning justified belief. We have just seen that a mere evaluative standard, even when applied to an evaluative standard, does not explain what constitutes the adequacy, or correctness, of an evaluative standard. The verificationism that implies otherwise offers only conceptual circularity – a kind of circularity inimical to effective explanation.

Our question about adequacy for the evaluative project leads thus far to a bothersome dilemma: either naivete or circularity. Each of these horns is

troublesome, if not fatally sharp, for effective philosophical explanation.

An analogous dilemma challenges the explanatory project. Suppose that I formulate and accept conditions that explain, in nonepistemic terms, what makes a belief justified. An explanatory standard for epistemically justified belief will characteristically take this form:

A belief is epistemically justified if (and only if) it satisfies explanatory conditions *C*.

Perhaps I am a “process-realist” whose conditions appeal to reliable, truth-conducive processes of belief-formation in the absence of defeat, processes such as reliable perception, memory, and introspection.³

Whatever my explanatory conditions are, this question will arise: In virtue of what, if anything, are my proposed conditions adequate, or correct, as a solution to the explanatory project, at least for me? In other words, what, if anything, constitutes the correctness, at least for me, of my explanatory conditions as an answer to the explanation-seeking question of what constitutes epistemic justification? As before, our question can take a modal form for cases of false epistemic standards: The issue then is what *can* constitute correctness of one’s standards.

We now return to a familiar theme. Either I take the previous question about adequacy seriously, or I do not. If I do not, I shall consider it negligible, as not needing an answer. In that case, my acceptance of my explanatory conditions for justification will be naive, or superficial. My acceptance will then lack a cogent reply to an intelligible, legitimate question about what, if anything, constitutes the adequacy of my explanatory conditions. This is a species of the objectionable meta-epistemic naivete mentioned before. If, alternatively, I do take the previous question seriously, I shall seek an answer. Because, by hypothesis, I already accept a set of explanatory conditions for justification, I might offer those conditions to explain what constitutes the adequacy of those conditions as a solution to the explanatory project. (I assume, as before, that an appeal to an infinite regress of epistemic standards fails.)

We may begin by asking whether any special analogous dilemma troubles the aforementioned semantic project: the project of specifying, in informative terms, what it means to say that something is epistemically justified. By “special

dilemma” I mean a dilemma peculiar to the semantic project. Suppose that I answer the semantic project by defining “epistemic justification” via talk of what is “permissible” relative to a specific set of epistemic rules or, alternatively, via talk of what is “good” from the standpoint of acquiring true beliefs and avoiding false beliefs.⁴ Would either such answer raise a distinctive dilemma for the semantic project, a special dilemma analogous to our previous dilemmas?

The following question might seem to raise an analogous dilemma: What does it mean to say that my definition of “epistemic justification” is itself epistemically justified? This question does not automatically generate a dilemma of naivete or circularity. Without being naive, one might reject the question as resting on a false assumption that it makes sense to say that definitions of “epistemic justification” are themselves epistemically justifiable. One might hold, without naivete, that definitions of “epistemic justification” are preconditions for talking about epistemic justification but are not themselves candidates for such justification. One’s reason might be that definitions are ultimately stipulative – even if not capriciously stipulative – and thus are neither true nor false independently of stipulation. (This does not imply that one cannot be epistemically justified in holding that one endorses a certain definition. The proposition that one endorses a certain definition is not the same as the definition itself.) Talk of epistemic justification, on this view, makes sense only relative to a conceptually prior notion of epistemic justification; and this prior notion is a conceptual precondition, but not a recipient, of epistemic justification. Such a view seems not to be “naive” in the way the initial response to our dilemmas was.⁵

One might, alternatively, allow for the epistemic justifiability of definitions of “epistemic justification,” but plead innocent to any objectionable circularity in answering the previous question. Suppose that we accept a definition of “epistemic justification” as what is permissible relative to a specific set of epistemic rules. When asked what it means (at least for us) to say that our definition is epistemically justified, we can plausibly reply that our definition itself gives the answer: namely, our accepting the definition is permissible relative to a specific set of epistemic rules. This seems unobjectionable. If the question asks just for what we mean by “our definition is epistemically justified,” we have given the full, most informative answer.

We cannot be plausibly accused of having given an answer that is circular or shallow.

Let us consider another question facing any answer to the semantic project:

What, if anything, constitutes the correctness (at least for myself) of my semantic standards for “epistemic justification” as an answer to the semantic project regarding what it means to say that something is epistemically justified?

The semantic project, as noted, seeks what it means to say that something (for example, a proposition or a belief) is epistemically justified.

If semantic meaning is person-relative, the semantic project aims to specify, in informative (or noncircular) terms, what it means for one to say that something is epistemically justified. (For now we can proceed with a loose notion of *informative* meaning as noncircular meaning, and a loose notion of *definition* as whatever specifies meaning for one.) Adequacy of a definition of “epistemic justification,” as an answer to the semantic project for myself, thus amounts to a specification, in informative terms, of what I actually mean by “epistemic justification.” The latter specification, being adequate, is just my actual informative definition of “epistemic justification.”

My actual informative definition of “epistemic justification,” even if multifaceted and somewhat vague, is definitive of my adequate, or correct, solution to the semantic project. It is what determines the adequacy – the correctness – of my solution. If I do not have any definition of “epistemic justification” (or some synonymous expression), I shall lack a correct solution to the semantic project. In that case, I shall lack a determinate notion of epistemic justification. (The definitions pertinent now need not be highly specific in the way typical analyses of notions of justification and knowledge are; on this score, we may contrast definitions and analyses.)

When a definition is adequate, or correct, as a solution to the semantic project, *it* is what explains the adequacy of the solution in question; it then amounts to what one means by “epistemic justification” (or some synonymous expression). There is thus an internal connection between (a) what one informatively means by “epistemic justification,” (b) one’s actual informative, noncircular definition of “epistemic justification,” and (c) one’s correct solution to the semantic project. Correctness here is constituted by what one actually means. What

one means by “epistemic justification” does not, however, preclude one’s understanding other notions of epistemic justification.

We might ask how one can know or justifiably believe what one’s actual definitions are, but this is not the concern of the question posed above. Its concern is rather to ask what constitutes adequacy of a solution to the semantic project; and its answer comes directly from considerations about what one informatively means in saying that something is epistemically justified. Insofar as one informatively means anything in saying that something is epistemically justified, the semantic project will not generate for one the dilemma of naivete or circularity posed earlier. One will then have a constitutive standard for a correct answer to the semantic project: an answer stating what it is that one informatively means in saying that something is epistemically justified. The semantic project thus lends itself to a cogent answer to question.

Semantic considerations hold the key to avoiding the other manifestations of the dilemma of naivete or circularity, and to solving the evaluative and explanatory projects. More specifically, considerations about what one informatively means by “(adequate for) discerning justified beliefs” and by “(adequate for) explaining justification” enable effective explanations for the evaluative and explanatory projects. What one means by “(adequate for) discerning, or evaluating, justified beliefs” can effectively explain what constitutes correctness, at least for oneself, for an answer to the evaluative project. Similarly, what one means by “(adequate for) explaining justification” can effectively explain what constitutes correctness, at least for oneself, for an answer to the explanatory project. In the same vein, what one means by “(adequate for) effectively explaining” can effectively explain what constitutes correctness, at least for oneself, for an answer to a project seeking effective explanation.

If we neglect the internal connection between what one informatively means by “evaluating justification” and “explaining justification” and one’s correct solution to the evaluative and explanatory projects, we risk changing the topic from what is actually one’s correct solution. We then risk introducing such other topics as what solution is *justified* and what solution is *someone else’s* correct solution. We saw before how verificationist approaches to correctness change the topic to justification and thereby risk conceptual circularity. We need instead a semantic approach to

Semantic Foundationalism

The following broad thesis acknowledges a central role for semantic considerations in matters epistemological:

One's explaining, evaluating, and arguing for epistemic justification and one's answers to questions about correct standards for such matters properly end in considerations about an operative *notion* for one regarding epistemic justification.

Call this view *semantic foundationalism*. It incorporates and extends the earlier lessons about semantic considerations that free the three epistemological projects from a dilemma of naivete or circularity.

Notions and conceptual commitments

One's *notion* of epistemic justification comprises semantic standards constitutive of what it is to be one's correct use of such terms as "epistemic justification," "epistemic support," and "epistemic warrant." (We may add any synonymous term to the list, including any synonymous term from some other language.) Such semantic standards are crucial to the individuation of notions, as they determine what a notion is a notion *of*; they determine the constitutive conditions of satisfaction, or fulfillment, of a notion.

I shall generally talk of "standards for correct use" rather than "standards for what it is to be correct use," but the two locutions are equivalent for me. Both concern the *category* of correct use for me, not just particular cases of correctness in use. One's semantic standards determine the category of correct use for one: that is, relative to what one means by "correctness in use" (or some synonymous phrase). The pertinent relativity is, then, semantic, not just a matter of mere belief about what is true.

Semantic standards contribute determinacy to what tokens mean for one. Linguistic meaning, as ordinarily understood, has some determinacy: It does not admit every conceivable interpretation or use of a token as correct. It excludes some interpretations or uses as incorrect; and the more specific interpretations or uses it excludes, the more determinate it is. Exclusion of interpreta-

correctness in terms of what one means; this approach can avoid the dilemma of naivete or circularity.

In adopting a semantic approach to correctness, we can avoid troublesome commitments to ontologically dubious truth-makers for semantic epistemic standards: for example, Platonic epistemological entities and equally questionable epistemological natural kinds.⁶ We can thereby avoid familiar skeptical questions about the support needed for claims to the existence of such truth-makers. If we can avoid such questions without loss, we are wise to do so.

Talk of *one's* correct solution to the semantic project is not automatically talk of the solution one believes to be correct. It rather is talk of what is correct relative to one's own understanding of what constitutes correctness: that is, relative to what one *means* in saying that something is correct. The relativity here is thus semantic, not simply doxastic. Understanding of what constitutes correctness can vary even if one's own notion of "a notion of correctness (generally characterized)" sets some limits on what in general is, for oneself, a notion of correctness. In accepting, offering, or considering a solution to one of the epistemological projects, a theorist is presupposing a notion of correctness: a notion of a correct solution. This presupposed notion can, and sometimes does, vary among theorists, with respect to its specific requirements.

Restricting my account to my own specific understanding of correctness would obviously rob the account of comprehensiveness. My account would then exclude many theorists who now fall within its explanatory scope. The semantic relativity I acknowledge does not preclude notions of objective correctness: correctness that does not depend on one's believing that it obtains. Notions of correctness and meaningful claims to correctness do depend on someone's understanding of correctness; and such notions and claims are involved in one's accepting, offering, or considering a solution to an epistemological project. It does not follow, however, that the existence of what makes a claim correct – apart from its being described as a determinant of correctness – depends on one's own understanding of correctness. Even if a notion of correctness depends on someone's understanding, a relation of correctness (on some notions) can be independent of conceivers' beliefs that it obtains. My assumption of semantic relativity fits with this consideration.

tions or uses comes from standards constitutive of correct use. Such standards guarantee that not just *any* arbitrary interpretation or use of a token is correct for one. An interpretation or a use of a token is incorrect for one if it violates one's semantic standards for correct use of that token; a requirement of correctness is thus conformity to one's semantic standards. In the absence of determinacy as exclusion, meaning languishes, indeed expires.

The determinacy of linguistic meaning as exclusion need not come from "objective modal facts" or any such ontologically questionable basis. It can come simply from one's commitment to interpret or to use a token in certain ways to the exclusion of other ways. Conceptual necessity need have no basis deeper than one's adopted constitutive standards for correct use of terms; it can arise from the exclusions due to one's semantic standards.

Semantic standards are not equivalent to empirical generalizations, but are presupposed by such generalizations. The subject- and predicate-terms of an empirical generalization require determinacy from standards for correctness in interpretation if that generalization is to be intelligible and empirically confirmable (as correct). Experienced features by themselves do not yield standards constitutive of (the category of) correctness in interpretation for one. Denial of an empirical generalization involving a notion of knowledge, for example, does not entail either meaninglessness or a change of subject (owing to a different notion at work). Denial of a semantic standard, in contrast, does entail either meaninglessness or a change of subject. Semantic standards are, then, irreducible to empirical generalizations.

An internal connection obtains between what a notion is *of*, for one, and that for which the notion sets the condition "what it is to be that thing," for one. One's notion of justification, for instance, sets constitutive, or defining, conditions for what it is to be justification, at least for oneself, and thereby sets one's constitutive conditions for genuine justification, in the relevant sense. (Recall that the relativity here is semantic, not merely doxastic.) A notion of epistemic justification, we shall see, is not itself a candidate for being justified by considerations independent of fixing meaning, or of conceptual commitment. We shall also see that certain notions can themselves have a kind of purpose-based, instrumental support.

A notion of justification is *operative* for one in a particular situation when one is committed to that

notion as specifying what constitutes justification (for one) in that situation. One's currently operative notion of justification need not be one's only understood notion of justification. I might understand various specific notions of justification while being committed to only one as specifying what actually constitutes justification (for me). I could recognize that the alternative specific notions are amplifications of a common notion of justification generally characterized; still, I would not regard all the alternative notions as specifying what actually constitutes justification (for me). I could, nonetheless, adopt a variety of specific notions of justification that specify different species of justification (for me): moral justification, legal justification, and so on.

Commitment to a notion of epistemic justification is just commitment to certain constitutive, semantic standards for one's correct use of such terms as "epistemic justification," "epistemic support," and "epistemic warrant." Call any such commitment an *operative conceptual commitment*. We can now accommodate such talk as: I understand what Jones means by "justification," but that is not what *I* mean by "justification"; or I understand Jones's specific notion of justification, but Jones's specific notion is different from my (operative) specific notion of justification. Such talk suggests that we should not confuse conditions for an operative notion and conditions for an understood notion.

Difference in conceptual commitments, like difference in analyses, entails a difference in specific notions at issue. People operating with different specific notions of *X* can still, however, understand and even intentionally use common notions of *X* *generally characterized*. We thus can talk intelligibly of various notions of some one thing (generally characterized).

Even if you and I understand and intentionally use a common unspecific notion of "epistemic justification generally characterized" (for example, as permissibility relative to a set of rules of a certain sort), we can still differ on the actual rules conferring justification and thus have different operative notions of justification at a level of specificity. Even so, I could understand your notion of justification and acknowledge it as your elaboration, different from mine, on our common notion of justification generally characterized. I could therefore regard your notion as a notion of justification even though it differed from my specific notion of justification. I could, then, under-

stand your specific notion of justification, even if it differed from my operative notion of justification at a level of specificity.

The relevant talk of *meaning* and of constitutive conditions *set by* one's operative notion of justification involves whatever constitutes, or actually makes up, one's operative notion of justification. One's operative notion of justification is the logical basis for one's "conceptual truths" about justification, those truths being whatever logically follows from one's operative notion. They are "one's" conceptual truths just in virtue of following logically from one's operative notion. (We might introduce a stricter notion of conceptual truth requiring one's awareness of a logical connection between the truth and one's notion, but I shall not.)

An operative notion (or definition) of justification need not be a logical analysis of justification in terms of specific nondisjunctive conditions that are individually necessary and jointly sufficient for justification. What one means by "justification" may be based on "exemplars": certain paradigm-cases of justification that do not yield a logical analysis in terms of specific nondisjunctive necessary (and sufficient) conditions for justification.⁷ Notions (and definitions) need not conform to rigid standards for an "ideal" or a "clean" logical analysis. They rather can depend on various exemplars to yield a standard that offers some determinacy by exclusion (cf. Wittgenstein 1953, sections 71, 75, 87).

What is constitutive of the meaning of a statement or term for one does not reduce to mere talk of logically necessary and sufficient conditions. What is merely sufficient for the truth of *P* for one is not necessarily part of the meaning of *P* for one; so also for what is merely necessary for the truth of *P*. Similarly, what is logically necessary and sufficient for the truth of *P* can include components extraneous to the meaning of *P* for one. Suppose that *C1* and *C2* constitute the meaning of *P* for one. In that case, the following will be logically necessary and sufficient for the truth of *P* for one: $(C1 \ \& \ C2) \ \& \ [(C1 \ \& \ C2) \ \vee \ R]$. The latter disjunction is, by hypothesis, not part of the meaning of *P* for one. Constituents of meaning for one are exhausted by the actual components of one's constitutive standards for correct use of terms.

One's operative notion of epistemic justification is neither correct nor incorrect for one apart from a determinate semantic standard for correctness of a notion of justification. Even if epistemic justifica-

tion is a universal or a natural kind, the correctness or incorrectness of a notion for one relative to that universal or natural kind requires a determinate semantic standard for correctness for one. Lacking such a standard, one will not have a determinate notion of correctness of a notion, and thus one will not understand anything specifying correctness of a notion. Correctness of a notion for one will then lack its needed foothold in what one understands. As for (possible) objective, conceiving-independent relations of correctness, our talking intelligibly about them, and justifiably affirming them, requires a determinate notion of correctness.

Issues of conceptual correctness can arise indirectly, relative either to the implications of an alternative notion of justification or to claims about the scope or application of an operative notion of justification. We can ask, for instance, whether it is correct that a particular operative notion of justification captures either the conditions for a certain common use of the term "justification" or the conditions for my operative concept or your operative concept of justification. An operative notion of justification thus can raise questions of correctness indirectly, in virtue of accompanying claims about scope or application.

A notion of justification, like any determinate notion, rests on a semantic standard that has broadly normative significance. At a minimum, a semantic standard sets a constitutive normative condition for correct application of certain terms.⁸ When a semantic standard is prescriptive (for example, "Let justification be such and such!"), it will not itself be true or false in the way assertions can be. Normative semantic standards can, however, give semantic significance to definitional assertions about justification. They can offer semantic norms, or constitutive standards, for correct use of epistemic terms and locutions in definitional assertions. Such standards can thereby contribute semantic significance to a truth-valued definitional assertion: for example, "Justification, by definition, is truth-conducive belief-formation in the absence of defeat."

A definitional assertion could be made true for me by my adopting a certain semantic standard (for example, a certain definitional prescription): that is, adopting the standard as constitutive of correct use of a term (for me). Such a definitional truth could arise for a person from its stating only what is prescribed by that person's operative definitional prescriptions. Semantic standards need not appeal to prior synonymy relations, nor need

they be unrevisable. They thus can escape worries from Quine (1951) about the epistemological significance of considerations of meaning.

Philosophers uneasy with talk of notions might prefer to substitute talk of *constitutive standards*, or *norms*, for correct use of certain terms. This substitution might seem to remove without loss a questionable semantic component from the account under development. It allows us to talk simply of constitutive standards for correct use of epistemic terms, without talk of “notions” of justification. This move is just stylistic or notational, however, because notions, on my account, are just constitutive standards for correct use of certain terms.

Talk of notions does enable us to answer the following sort of question with ease: Why do you use those constitutive standards for correct use as constraints on *epistemic justification*? Such a question apparently presumes a distinction between one’s operative constitutive standards for correct use of such a term as “epistemic justification” and one’s operative notion of epistemic justification. If we regard the relevant operative semantic standards as *constitutive* of one’s operative notion of justification, we foreclose any such question. Talk of operative notions of justification thus serves a definite purpose in answering certain potentially troublesome questions about constitutive standards for correct use of epistemic terms.

Suppose that I am pursuing the explanatory project of explaining what constitutes justification. I endorse, we may assume again, the reliabilist view that justification consists in the truth-conduciveness of belief-forming processes, such as perception, in the absence of defeat. Being uneasy with reliabilism, you raise this now familiar question for my approach to the explanatory project: In virtue of what, if anything, are my reliabilist explanatory standards correct as a solution to the explanatory project, at least for myself? In reply, I might note that my acceptance of my reliabilist explanatory standards is itself supported by a truth-conducive belief-forming process in the absence of defeat. Thoroughgoing reliabilism requires this answer to a question about the justification of reliabilist standards. You, of course, will not be satisfied by such an answer because you have asked about constitutive conditions for the correctness of my standards, not their justification. You naturally want a more suitable answer. A true statement, of course, is not necessarily a correct answer to a question at hand.

The best answer available to typical “process-reliabilists” is that their specific semantic standards for “correct explanation of epistemic justification” involve considerations about reliable processes of belief-formation in the absence of defeat. These standards involving reliable processes are not, for typical process-reliabilists, extrinsically related to what it is correctly to explain epistemic justification; they rather are internally, or conceptually, related in virtue of an operative specific notion of correctly explaining epistemic justification.

An indicator of the conceptual nature of the relation between reliability and explaining justification for typical process-reliabilists is that they exclude the possibility of a mistake in their assumption that epistemic justification is correctly explained by appeal to reliability of belief-forming processes. They simply are not genuinely open to the possibility of a counterexample to this fundamental assumption of their reliabilism, even if other tenets of their theory admit of falsifiability. The fundamental assumption in question is, then, a conceptual truth for typical process-reliabilists: It is actually part (or at least an implication) of their operative specific notion of epistemic justification that beliefs resulting from truth-conducive processes in the absence of defeat are epistemically justified.

If semantic considerations figure crucially in what it is to be justification for a person (where the relativity is semantic), they should have some representation in an informative explanation of what it is for a person to explain, evaluate, or argue for justification. Semantic foundationalism accommodates this lesson by acknowledging a crucial role for considerations about an operative notion for one regarding epistemic justification. Explaining, evaluating, and arguing for justification come to an end, according to semantic foundationalism, with an appeal to an operative notion regarding justification.⁹ We need not pursue a detailed account of meaning and notions now, as semantic foundationalism remains neutral on most of the controversial issues about the exact conditions for fixing meaning and having a notion.

Arguing for justification

Semantic foundationalism entails that arguing deductively for justification for any statement – including any statement about justification – con-

sists paradigmatically of giving an argument in accord with, or reducible to, the following schema:

1. By my conceptual commitment, justification (for either my explanatory or evaluative epistemic standards, for example) consists in conditions *C* (for example, truth-conducive belief-formation in the absence of defeat).
2. The statement that *P* (expressing, for example, my explanatory or evaluative epistemic standards) satisfies *C*.
3. Hence, the statement that *P* is justified (at least for me).

This is an argument schema for one's deductively arguing for justification for a statement: 'that is, one's arguing deductively that a statement is justified (at least for oneself), not just that it is true. When the instances of steps 1 and 2 are true, the argument 1 through 3 for the justification of *P* will be sound; otherwise, we shall have an unsound deductive argument for the justification of *P*.

Schema 1 through 3 does not specify conditions for one's merely having justification for a statement. Merely having justification for a statement, on a common understanding, does not require giving an argument at all. It requires only one's having undefeated evidence for the statement in question; it does not require even one's having a notion of justification. In arguing or merely claiming that someone has justification for a statement, however, a theorist must employ a notion of justification. So long as we engage in any project of epistemology, by making claims about justification and knowledge, we shall confront the lessons of semantic foundationalism.

Semantic foundationalism recommends the use of arguments instantiating 1 through 3 to answer challenges to the justification of, among other things, one's answers to the explanatory and evaluative projects. Why? The answer is straightforward: What it is to be a justification, for one (where the relativity is semantic), depends on – indeed, is constituted by – one's operative notion of justification, and arguments instantiating 1 through 3 will conclusively ground justification in the notion that is one's semantic foundation. The pertinent semantic foundation defines what it means for one to say that something is justified, not what it means for one to say that we may regard something as justified. Schema 1 through 3 acknowledges the crucial role of an operative

notion of justification, as a semantic foundation, in what justification is for one.

When the relevant semantic foundation captures what one actually means in saying that something is justified, it enables an argument for justification that is sound, or correct, relative to one's actual operative semantic standards. In that case, an argument exemplifying 1 through 3 will be above reproach with respect to giving sound justification – at least for the person in question. Merely cogent justification, as just observed, can follow suit, but without the soundness: Justification can be deficient from the standpoint of correctness. We must distinguish, in any case, the justification of the statement justified according to the conclusion of 1 through 3 (namely, *P*) and the justification of 3 itself. Lower-order justification can get by without higher-order justification, despite any bothersome temptation to level-confusions.

Evaluating and explaining justification

Schema 1 through 3 illustrates how arguing deductively for justification for any statement, including any statement that ascribes justification, can properly end in considerations about an operative notion regarding justification. What about the evaluating and explaining of justification?

Semantic foundationalism extends readily to arguments of the following forms:

By my conceptual commitment, correct evaluation, or discernment, of epistemically justified belief consists in conditions *C*; evaluative epistemic standard *E* requires just *C*; hence, *E* is a correct evaluative epistemic standard, at least for me.

By my conceptual commitment, correct explanation of epistemic justification consists in conditions *C*; explanatory epistemic standard *E* requires just *C*; hence, *E* is a correct explanatory epistemic standard, at least for me.

These deductive arguments deliver conclusions about what evaluative and explanatory epistemic standards are correct. They do so on the basis of considerations about one's operative notions of correctly evaluating and correctly explaining epistemic justification. Schema 1 through 3, in contrast, illustrates how we can argue deductively, and fully, for conclusions affirming the *justification* of evaluative and explanatory standards. Arguing for

correctness of epistemic standards is one thing; arguing for their justification, another.

The lessons of semantic foundationalism, as we should expect, apply to semantic foundationalism itself. In particular, what one means by “properly end” in the earlier summary statements of the view will play a definitive role in what it means for semantic foundationalism to be a correct view. In fact, if we understand “properly end” to connote “correctly end,” we may treat semantic foundationalism as conceptually correct in virtue of semantic considerations regarding the explaining, evaluating, and arguing for justification and the correctness of corresponding epistemic standards for such matters.

Worries about circularity are misplaced now, as section 2 suggested in connection with the semantic project. Semantic foundationalism can be true in virtue of semantic considerations; and when asked what it means to say that semantic foundationalism is true, we can – indeed, must – appeal to the pertinent semantic considerations. After all, conceptual correctness is, by definition, correctness in virtue of meaning.

Going beyond mere correctness, schema 1 through 3 enables a full justification, even a sound justification, of semantic foundationalism for anyone with suitable conceptual commitments regarding justification. Consider an explanationist notion of justification (broadly in accord with Moser 1989, pp. 260–5) favoring an epistemological view that explains what needs to be explained better than any available competitor, while raising no decisive problems in doing so. We may regard explanation as the answering of pertinent why-, what-, and how-questions; and we may regard what needs to be explained as the (generally characterized) explaining, evaluating, and arguing for justification and the standards for correctness for such matters.

Betterness of explanation is a function of the problems raised and problems avoided by an explanation in comparison with competing explanations; but individual theorists can, and sometimes do, wield different standards for what exactly constitutes a *problem* for an explanation. An explanation that beats, or is as good as, any available competitor for one is a best explanation for one. An explanationist strategy for justification often takes this line: Let us consider the problem-questions (say, about justification) in *this* way, to see if we can answer the questions in a less problematic way than the available competitors; if we can,

justification accrues accordingly. Even here, as semantic foundationalism suggests, an operative, explanationist *notion* of justification is crucial.

Semantic foundationalism owes its correctness, then, to semantic considerations, but can derive justification from explanatory considerations. Conceptual truths, holding in virtue of conceptual commitments, can have explanatory value; that is, truths arising just from what one means can contribute to the answering of explanation-seeking why-, what-, and how-questions. When conceptual truths do have explanatory value, they can accrue justificational value accordingly (at least relative to an explanationist notion of justification of the general sort characterized in Moser 1989). An explanationist approach to justification is especially suited to bear on what philosophers typically do: offer answers to explanation-seeking questions.

4 Objections and Metaphilosophical Lessons

Notions and philosophical problems

Philosophers often decide on their philosophical notions and their conceptual systems relative to their own purposes, or ends, in theorizing.¹⁰ Philosophers, and not denizens of an impersonal external world, are doing their conceiving relative to their purposes in theorizing. An external world seems not to set by itself the boundaries of our philosophical concepts for us. We evidently set the constraints we do in light of a wide range of differing purposes. The simple ordering of “features” in perceptual experience is only one among many possible ends that can constrain concept-formation.

Notions, as section 3 suggests, comprise semantic standards constitutive of the correct use of terms. Many contemporary epistemologists adopt and use notions of justification that share a core with the notions of justification assumed by various leading epistemologists from Plato to Kant to Russell. The core of these notions derives from an interest in what Plato, Kant, and Russell were regarding as the third condition for knowledge. This common core is rather thin in certain cases, including in some cases only sketchy standards involving vague conditions of “well-foundedness,” “nonaccidental connectedness with the truth condition,” “likelihood of truth,” and so on. This common core does, nonetheless, set some general

constraints. It rules out, for instance, justification as obviously contradictory belief-formation and, more generally, justification as an “anything goes” policy of belief-formation.

I may hold that another person’s semantic standards for “justification” must share the core of my own acknowledged set of such standards to qualify as semantic standards regarding justification. What exactly constitutes the relevant core will be up to an individual theorist. My own set of semantic standards for “justification,” on this view, will not conflict with the core of any alternative set that I (properly) call “semantic standards regarding *epistemic justification*.” My semantic standards are thus my anchor not only for arguing for justification but also for deciding on genuine concepts of justification. I decide what counts as a notion of justification from my conceptual perspective: in particular, from the perspective of my adopted conceptual core (necessary) for any concept of justification. My conceptual perspective *might* be shared by others, but we have no conclusive reason to think that it must be shared – even if it must be in principle understandable, or shareable, by others.¹¹

An important methodological lesson emerges here. Semantic foundationalism suggests that the perennial nonempirical problems of epistemology – and of philosophy in general – perplex us endlessly and even seem insoluble, typically because those problems are formulated via insufficiently explicit or detailed semantic standards for their key terms. The epistemological problems in question include: Is knowledge justified true belief? Does empirical justification have foundations? Is epistemic justification a function of reliable belief-formation? Once the relevant semantic standards for such terms as “knowledge” and “justification” are given sufficient detail (and this we typically do with an eye to agreement with our antecedent semantic standards and semantically based adequacy conditions for resolving problems of interest), then the problems surrender. They vanish as a result of our expanded, more specific semantic foundations. One may still face problems concerning which epistemological notions actually best serve one’s theoretical ends, but that is a practical problem, a problem of applying relevant standards.

When the needed semantic standards lack specificity, perennial philosophical controversy ensues, owing typically to varying implicit specifications of those standards by different philosophers. Such is the usual way of “conflict” in traditional epistemology and in philosophy gener-

ally. The conflict here is actually “conflict,” because variation in implicit specifications of semantic standards entails that philosophers are really talking about different matters. The resolution of such “conflict,” in typical cases, requires *explicit* formulation of the notions formulating the “conflict.” It requires the making explicit of semantic foundations. This will open the door, the only door, to our resolving the problem under dispute – if only by showing that the participants to the “conflict” are really talking about different matters.

Some philosophers will worry about specifying, or amplifying, a vague semantic standard in the *right* way. Such a worry seeks constraints for the “correct” amplification of a notion. We may think of notions as semantic packages, of varying complexity, constituted by semantic standards for the correct use of terms. We may think, for example, of the traditional philosophical notion of knowledge, stemming from *Theaetetus* 202c, as a semantic package including concepts of justification, truth, belief, and the absence of defeaters. Semantic packaging, I have suggested, is done by conceivers, even *if* the extension of a concept is sometimes conceiver-independent.¹²

Much semantic information comes to us seemingly “prepackaged,” via social inheritance. Some inherited concepts, however, need our special repackaging: for example, amplifying for our philosophical purposes. Some concepts are too vague or too sketchy to settle our philosophical concerns. We thus need to fill in or to refill the semantic standards accompanying such concepts. Given strict conditions for individuating concepts, this will entail the formulation of a different concept, but we may talk of different specifications of a concept *generally characterized*. It is misleading to think of the task of supplementing or revising semantic standards as analogous to empirical discovery in the natural sciences. A better model is that of *construction* relative to theoretical purposes and accepted adequacy conditions for meeting those purposes. Even here, however, a *notion* of adequate, or correct, construction will play a crucial role in explaining, evaluating, and justifying alteration of semantic standards.

Worries about correctness of conceptual amplification may stem from one’s acceptance of certain adequacy conditions for solving philosophical problems relevant to the concept being amplified. We thus should beware of regarding our preferred constraints on such correctness as being categorically

binding. A constraint on semantic packaging, or on conceptual amplification, is either internal or external to a semantic package under revision. If *internal*, a constraint comes solely from semantic entailment relations with antecedent components of the package, and thus from prior conceptual commitments underlying that package. Such a constraint will be semantically entailed by an antecedent component of the package under revision. If *external*, a constraint will not be semantically entailed by an antecedent component of the package.

Use of an internal constraint simply makes explicit what is already implicit in a semantic package. We might, for example, come to see that our concept of knowledge involves a notion of adequate evidence, upon coming to see that our concept of knowledge involves an internalist notion of perspectival, person-relative justification. This would be a moderate internal amplification. Use of an external constraint is potentially more revisionary. Such a constraint comes from an independent adequacy condition accepted by a person assessing conceptual amplification. Independent adequacy conditions for correct formulation of concepts can, and do, vary among philosophers. This partly explains why many substantive philosophical disagreements resist resolution.

Familiar examples of independent adequacy conditions include (a) what one "would say" under certain imagined conditions, (b) some independent pattern of language use, for example, ordinary language use, and (c) certain explanatory requirements for an adequate resolution of a problem, for example, simplicity and comprehensiveness. Such constraints typically are not internal to concepts under amplification. They rather are introduced by a theorist wielding a particular *notion* of "correctness" or "justification" as a standard for conceptual amplification. A theorist thus may judge certain semantic standards as correct or incorrect, on the basis of such constraints. Such a judgment will rest for its intelligibility, however, on certain other semantic standards for correctness or incorrectness: at least on a semantic standard for correct conceptual amplification. We again see a particular instance of a recurring lesson from semantic foundationalism.

Conceptual Relativism Introduced

Semantic foundationalism allows for a kind of semantic, or conceptual, relativism. Different phi-

losophers can, and in some cases do, have different operative semantic standards for the use of the term "justification," at least at a level of specificity. This consideration should be underwhelming; indeed, it is a platitude if anything is. If this is all one means in saying that philosophers can, and in some cases do, have different operative concepts of justification, we have nothing to dispute.

If operative concepts are individuated by one's pertinent operative semantic standards, we need to allow for substantial variability of operative concepts. The contrary view entails the epistemological myth of the definite article: the view that there must be, or at least is, such a thing as *the* concept of justification or knowledge. Even so, different people can, and in some cases do, share a notion of justification *generally characterized*. They can, moreover, typically understand alternative specific notions of justification, even if they do not understand them as specifying what justification actually consists in. One's own specific notion of justification defines what justification consists in (at least for oneself).

Conceptual relativism does not entail *substantive relativism*: the view that whatever one takes to be correct, right, or justified is actually correct, right, or justified. On some concepts of epistemic and moral justification, fundamental epistemic and moral requirements show no cultural or personal variability, but rather are universal in scope. The rejection of substantive relativism thus leaves conceptual, or semantic, relativism untouched. The conceptual relativism allowed by semantic foundationalism does not entail an "anything goes," or substantively relativistic, attitude toward epistemic assessment.

Conceptual relativism still might seem bothersome. It *apparently* implies that conceptual variability admits of no rational assessment. Suppose a student, Jones, comes to us for advice on how to regulate beliefs. Jones is torn between two notions of justification. One notion, from a thoroughgoing mystic, implies that Jones can justifiably believe whatever arises spontaneously, without interference from inference. A second notion, from a W. K. Clifford-style scientist, implies that Jones can justifiably believe only what enjoys evidence in accord with an experimental, scientific method. Jones, being torn, asks us what she should believe. Should she believe in accord with the mystic's notion or the scientist's notion?

We naturally reply by complicating Jones's options. We introduce some additional notions of

justification in circulation. If in doubt, we advise, then complicate. Where, however, does this advice take Jones? She will certainly wonder which of the various notions her believing should (aim to) follow. Such wondering, nonetheless, is thus far seriously indeterminate, and can be seen to be so upon reflection.

We do not yet know the meaning – the semantic standards – of “should” in the question of what Jones *should* believe. This question becomes manageable determinate only when the meaning of “should” is made definite: only with the specification of semantic standards for “should.” Such standards will link the relevant use of “should” to determine notions of *obligation*, *permissibility*, and *justification* – to determinate semantic requirements constituting relevant notions of obligation, permissibility, and justification. Once this specification – this linkage – obtains, we have the needed basis for understanding Jones’s question and for seeking an answer. We then understand what notion of obligation is determining Jones’s use of “should.”

If, for example, Jones is after what she should prudentially believe (instead of what she epistemi-

cally or morally should believe, for example), we shall ask what notion of justification is conducive to prudential belief in the relevant sense – a sense that may itself need specification.¹³ The only remaining problem is practical, a problem of applying pertinent standards. (Jones may, of course, be asking something indeterminate for herself; in that case, we shall not have a determinate question to answer.) This recourse to an operative notion, to what one means, obviously fits with semantic foundationalism. It ultimately resolves a typical normative philosophical question by semantic considerations.

Semantic foundationalism clearly bears on nonepistemic and epistemic species of justification. If one seeks to give an instrumental justification for something, this justification will rest on considerations about a notion of instrumental justification relative to which certain factors can and do provide such justification. Corresponding points hold for moral, aesthetic, and legal justification, for example. Semantic foundationalism thus ranges widely, over any domain of justification.

Notes

- 1 Some reliabilists, following Alvin Goldman (1979, pp. 602–3; 1980, pp. 28–9), treat the explanatory project as involving complex causal and historical considerations that preclude applicability for evaluation.
- 2 I shall not pursue the exact reason for the latter point, as it would take us too far afield. For pertinent discussion on the analogous topic of endless regresses of justification, see Cornman (1980, pp. 135–8), Post (1987, pp. 87–91), and Moser (1985, pp. 107–15; 1989, pp. 56–60).
- 3 As this is only a rough illustration, we may overlook the difficult question, for reliabilists, of what actual or counterfactual situations most exemplify the truth-conduciveness of belief-forming processes that confer justification. On that question, see Pollock (1984), Feldman (1985), and Moser (1989, pp. 196–202). For attempts at an answer, see Goldman (1988), Sosa (1991, pp. 281–4), and Schmitt (1992, ch. 6).
- 4 The former definition, made prominent by Roderick Chisholm, has been endorsed by Pollock (1986, pp. 7–8) and Alvin Goldman (1986, p. 59), among others; the latter, by Alston (1989, ch. 4).
- 5 Such a view seems to be suggested, if only vaguely, at various places in Wittgenstein’s *On Certainty* (1969, sections 608–20). The following two sections return to this view.
- 6 Cf. Alvin Goldman (1989, p. 143): “Whatever one thinks about justice or consciousness as possible natural kinds, it is dubious that knowledge or justificational status are [*sic*] natural kinds.”
- 7 On the role of exemplars in meaning, see Rosch (1978, 1983), Rosch and Mervis (1975), and Lakoff (1987, pp. 39–57; 1989).
- 8 We may now leave open the question whether the relevant terms might occur only in a language of thought, rather than in a socially shared natural language.
- 9 We shall see in the next subsection that semantic foundationalism applies to operative principles of inductive and deductive logic as well as to other justificatory standards. Semantic foundationalism thus fits with the suggestion of Feigl (1950, p. 118) that “we may say that the *rules* of logic in their totality [that are operative for us] *define what we mean by correct reasoning*.”
- 10 Here I agree with C. I. Lewis (1926, 1929). Cf. Waismann (1939) and Carruthers (1987).
- 11 For evidence that Wittgenstein’s influential views on privacy are no threat here, see Moser (1991, 1992) and Baker and Hacker (1985, pp. 173–9).

12 Semantic foundationalism is logically compatible with, but does not entail, a sort of realism about the extension of concepts. For some discussion of such realism, see Anscombe (1976) and Hacker (1986, ch. 11). My own commitment to semantic

foundationalism is, of course, coupled with conditional ontological agnosticism.

13 For elaboration on this point in connection with the semantics of obligation-talk, see Moser (1985, ch. 6).

References

- Alston, William, 1989. *Epistemic Justification* (Ithaca, NY: Cornell University Press).
- Anscombe, G. E. M. 1976. "The Question of Linguistic Idealism," *Acta Philosophica Fennica* 28, pp. 209–25; reprinted in Anscombe, *From Parmenides to Wittgenstein* (Oxford: Basil Blackwell, 1981).
- Baker, Gordon, and P. M. S. Hacker, 1985. *Wittgenstein: Rules, Grammar, and Necessity* (Oxford: Basil Blackwell).
- Carruthers, Peter, 1987. "Conceptual Pragmatism," *Synthese* 73, pp. 205–24.
- Cornman, James, 1980. *Skepticism, Justification, and Explanation* (Dordrecht: Reidel).
- Dummett, Michael, 1991. *The Logical Basis of Metaphysics* (Cambridge, MA: Harvard University Press).
- Feigl, Herbert, 1950. "De Principiis Non Disputandum . . ." in Max Black (ed.), *Philosophical Analysis* (Ithaca, NY: Cornell University Press).
- Feldman, Richard, 1985. "Reliability and Justification," *The Monist* 68, pp. 159–73.
- Goldman, Alvin I., 1979. "What Is Justified Belief?" this vol., ch. 27.
- , 1980. "The Internalist Conception of Justification," in P. A. French et. al. (eds), *Midwest Studies in Philosophy*, vol. 5 (Minneapolis: University of Minnesota Press).
- , 1986. *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- , 1988. "Strong and Weak Justification," in James Tomberlin (ed.), *Philosophical Perspectives*, vol. 2: *Epistemology* (Atascadero, CA: Ridgeview).
- , 1989. "Psychology and Philosophical Analysis," *Proceedings of the Aristotelian Society* 89, pp. 195–209; reprinted in Goldman, *Liaisons*, pp. 143–53 (Cambridge, MA: MIT Press, 1992). Reference is to this reprint.
- Hacker, P. M. S., 1986. *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, 2nd edn (Oxford: Clarendon Press).
- Lakoff, George, 1987. *Women, Fire, and Dangerous Things* (Chicago: University of Chicago Press).
- , 1989. "Some Empirical Results about the Nature of Concepts," *Mind and Language* 4, pp. 103–29.
- Lewis, C. I., 1926. "The Pragmatic Element in Knowledge." *University of California Publications in Philosophy* 6, pp. 205–27; reprinted in P. K. Moser and Arnold vander Nat (eds), *Human Knowledge*, pp. 201–11 (New York: Oxford University Press, 1987).
- , 1929. *Mind and the World-Order* (New York: Charles Scribner).
- Moser, Paul. 1985. *Empirical Justification* (Dordrecht: Reidel).
- , 1989. *Knowledge and Evidence*. Cambridge: Cambridge University Press.
- , 1990. "Some Recent Work in Epistemology," *Philosophical Papers* 19, pp. 75–98.
- , 1991. "Malcolm on Wittgenstein on Rules," *Philosophy* 66, pp. 101–5.
- , 1992. "Beyond the Private Language Argument," *Metaphilosophy* 23, pp. 77–89.
- Pollock, John, 1984. "Reliability and Justified Belief." *Canadian Journal Of Philosophy* 14, pp. 103–14.
- , 1986. *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield).
- Post, John, 1987. *Faces of Existence: An Essay in Non-reductive Metaphysics* (Ithaca, NY: Cornell University Press).
- Quine, W. V., 1951. "Two Dogmas of Empiricism," in his *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953).
- Rosch, Eleanor, 1978. "Principles of Categorization," in E. Rosch and B. Lloyd (eds), *Cognition and Categorization* (Hillsdale, NJ: Lawrence Erlbaum Associates).
- , 1983. "Prototype Classification and Logical Classification: The Two Systems," in E. Scholnick (ed.), *New Trends in Cognitive Representation* (Hillsdale, NJ: Lawrence Erlbaum Associates).
- Rosch, Eleanor, and Carolyn Mervis, 1975. "Family Resemblances: Studies in the Internal Structure of Categories," *Cognitive Psychology* 7, pp. 573–605.
- Schmitt, Frederick, 1992. *Knowledge and Belief*. London: Routledge.
- Sosa, Ernest, 1991. *Knowledge in Perspective* (Cambridge: Cambridge University Press).
- Waismann, Friedrich, 1939. "What Is Logical Analysis?" in his *Philosophical Papers*, ed. B. F. McGuinness (Dordrecht: Reidel, 1977).
- Wittgenstein, Ludwig, 1953. *Philosophical Investigations*, ed. G. E. M. Anscombe and R. Rhees, trans. G. E. M. Anscombe (Oxford: Basil Blackwell).
- , 1969. *On Certainty*, ed. G. E. M. Anscombe and G. H. von Wright, trans Denis Paul and G. E. M. Anscombe (Oxford: Basil Blackwell).

Reflective Equilibrium, Analytic Epistemology, and the Problem of Cognitive Diversity

Stephen Stich

This is a paper about different ways of thinking or cognitive diversity, as I shall sometimes say – and the problem of choosing among them. In the pages to follow I will defend a pair of claims. The first is that one influential proposal for solving the problem of cognitive diversity, a proposal that invokes the notion of reflective equilibrium, will not work. The second is much more radical. What I propose to argue is that although some of the objections to the reflective equilibrium solution turn on details of that idea, the most serious objection generalizes into an argument against an entire epistemological tradition – the tradition that I shall call “analytic epistemology.” Before attending to either of these claims, however, I will have to say something about how I conceive of cognition and cognitive diversity.

1 Cognition and Cognitive Diversity

Let me begin with a simplifying assumption that I hope you will not find wildly implausible. I shall assume that in humans and other higher animals there is a distinct category of mental states whose function it is to store information about the world. When the organisms in question are normal, adult humans in a culture not too remote from our own, folk psychology labels these states *beliefs*. Whether or not this folk label can be used appropriately for

the belief-like states of animals, automata, young children and exotic folk is a question of considerable controversy.¹ For present purposes, however, it is a controversy best avoided. Thus I propose to adopt the term “cognitive state” as a broad cover term whose extension includes not only beliefs properly so-called, but also the belief-like information-storing mental states of animals, young children and those adult humans, if any there be, whose cognitive lives differ substantially from our own.

Our beliefs, and the cognitive states of other creatures, are in a constant state of flux. New ones are added and old ones removed as the result of perception, and as a result of various processes in which cognitive states interact with each other. In familiar cases, folk psychology provides us with labels like “thinking” and “reasoning” for these processes, though once again the propriety of these labels becomes controversial when the cognitive states being modified are those of children, animals or exotic folk. So I will use the term “cognitive processes” as a cover term whose extension includes our own reasoning processes, the up-dating of our beliefs as the result of perception, and the more or less similar processes that occur in other organisms.

Cognitive processes are biological processes; they are something that brains do. And, like other biological processes, they have been shaped by natural selection. Thus it is to be expected that our genes exert an important influence on the sorts of cognitive processes we have. It is also to be expected that the cognitive processes of other

Originally published in *Synthese* 74 (1988), pp. 391–413; reprinted with kind permission from Kluwer Academic Publishers.

species with other needs and other natural environments will be in varying degrees different from those to be found among humans. But from the fact that genes inevitably exert a major influence on cognitive processes it does *not* follow that all of our cognitive processes are innate, or, indeed, that any of them are.

To see the point, we need only reflect on the case of language. My ability to speak English is a biological ability; processing English is something my brain does. Moreover, my genes are surely heavily implicated in the explanation of how I came to have a brain that could process English. Still, English is not innate. The ability to process English is an ability I acquired, and had I been raised in a different environment I might have acquired instead the ability to speak Korean or Lapp. This is not to deny that *something* relevant to language is innate. All normal human children have the ability to acquire the language spoken around them. And that is a very special ability. There is no serious evidence indicating that members of any other species can acquire human languages or anything much like them.

Now the point I want to stress is that, as far as we know, human cognitive processes may be like human language processing abilities. They may be acquired in ways that are deeply dependent on environmental variables, and they may differ quite radically from one individual or culture to another. Of course, it is also possible that human cognitive processes are much less plastic and much less under the influence of environmental variables. It is possible that cognition is more similar to digestion than to language. To make matters a bit messier, there is no reason a priori for all cognitive processes to be at the same point on this continuum. It may be that some of our cognitive processes are shared by all normal humans, while others are a part of our cultural heritage.² I am inclined to think that this last possibility is the most plausible one in the light of available evidence, and for the remainder of this paper I will take it for granted. But it must be admitted that the evidence is both fragmentary and very difficult to interpret.³

If we suppose that there is a fair amount of acquired diversity in human cognitive processes, and that patterns of reasoning or cognitive processing are to some substantial degree molded by cultural influences, it adds a certain urgency to one of the more venerable questions of epistemology. For if there are lots of different ways in which the human mind/brain can go about ordering and

reordering its cognitive states, if different cultures could or do go about the business of reasoning in very different ways, *which of these ways should we use?* Which cognitive processes are the *good* ones? It is just here that the analogy with language breaks down in an illuminating way. Most of us are inclined to think that, at least to a first approximation, one language is as good as another. The one you should use is the one spoken and understood by the people around you.⁴ By contrast, most of us are *not* inclined to accept this sort of thoroughgoing relativism about cognitive processes. If primitive tribesmen or pre-modern scientists or our own descendants think in ways that are quite different from the ways we think, few of us would be inclined to suggest that all of these ways are equally good. Some ways of going about the business of belief revision are better than others. But just what is it that makes one system of cognitive processes better than another, and how are we to tell which system of reasoning is best? In the remaining sections of this paper I want to consider one influential answer to this question. I shall argue that both the answer itself and the philosophical tradition it grows out of should be rejected.

2 Reflective Equilibrium as a Criterion for Assessing Cognitive Processes

The answer I will disparage was first suggested about three decades ago when, in one of the more influential passages of twentieth century philosophy, Nelson Goodman described a process of bringing judgments about particular inferences and about general principles of inference into accord with one another. In the accord thus achieved, Goodman maintained, lay all the justification needed, and all the justification possible for the inferential principles that emerged. Other writers, most notably John Rawls, have adopted a modified version of Goodman's process as a procedure for justifying moral principles and moral judgments. To Rawls, too, we owe the term "reflective equilibrium" which has been widely used to characterize a system of principles and judgments that have been brought into coherence with one another in the way that Goodman describes.⁵

It is hard to imagine the notion of reflective equilibrium explained more eloquently than Goodman himself explains it.

How do we justify a *deduction*? Plainly by showing that it conforms with the general rules of deductive inference. An argument that so conforms is justified or valid, even if its conclusion happens to be false. An argument that violates a rule is fallacious even if its conclusion happens to be true. . . . Analogously, the basic task in justifying an inductive inference is to show that it conforms to the general rules of *induction*. . . .

Yet, of course, the rules themselves must ultimately be justified. The validity of a deduction depends not upon conformity to any purely arbitrary rules we may contrive, but upon conformity with valid rules. When we speak of *the* rules of inference we mean the valid rules – or better, *some* valid rules, since there may be alternative sets of equally valid rules. But how is the validity of rules to be determined? Here . . . we encounter philosophers who insist that these rules follow from some self-evident axiom, and others who try to show that the rules are grounded in the very nature of the human mind. I think the answer lies much nearer to the surface. Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields unacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences.

This looks flagrantly circular. I have said that deductive inferences are justified by their conformity to valid general rules, and that general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement thus achieved lies the only justification needed for either.

All this applies equally well to induction. An inductive inference, too, is justified by conformity to general rules, and a general rule by conformity to accepted inductive inferences.⁶

There are three points in this passage that demand a bit of interpretation. First, Goodman claims to

be explaining what justifies deductive and inductive inferences. However, it is not clear that, as he uses the term, *inference* is a cognitive process. It is possible to read Goodman as offering an account of the justification of principles of logic and of steps in logical derivations. Read in this way, Goodman's account of justification would be of no help in dealing with the problem of cognitive diversity unless it was supplemented with a suitable theory about the relation between logic and good reasoning. But as several authors have lately noted, that relation is much less obvious than one might suppose.⁷ It is also possible to read Goodman as speaking directly to the question of how we should go about the business of reasoning,⁸ and offering a solution to the problem of cognitive diversity. This is the reading I propose to adopt.

A second point that needs some elaboration is just what status Goodman would claim for the reflective equilibrium test he describes. It is clear Goodman thinks we can conclude that a system of inferential rules is justified if it passes the reflective equilibrium test. But it is not clear *why* we can conclude this. Two different sorts of answers are possible. According to one answer, the reflective equilibrium test is *constitutive* of justification or validity. For a system of inferential rules to be justified just *is* for them to be in reflective equilibrium. Another sort of answer is that if a set of inferential principles passes the reflective equilibrium test, this counts as good *evidence* for them being valid or justified. But, on this second view, being in reflective equilibrium and being justified are quite different. One is not to be identified with the other. I am inclined to think that it is the former, constitutive, view that best captures Goodman's intentions. But since my concern is to criticize a view and not an author, I don't propose to argue the point. Rather, I will simply stipulate that the constitutive reading is the one I'm stalking.⁹

The third point of interpretation concerns the status of the claim that reflective equilibrium is constitutive of justification. On this point, there are at least three views worth mentioning. The first is that the claim is a *conceptual truth* – that it follows from the meaning of “justification” or from the analysis of the concept of justification. Like other conceptual truths, it is both necessarily true and knowable a priori. If we adopt this view, the status of the claim that reflective equilibrium is constitutive of justification would be akin to the status of the claim that being a closed, three-sided

plane figure is constitutive of being a triangle, though the claim about justification is, of course, a much less obvious conceptual truth. A second view is that the claim is a non-conceptual necessary truth that is knowable only a posteriori. This would accord it much the same status that some philosophers accord to the claim water is H₂O. Finally, it might be urged that the claim is being offered as a stipulative proposal. It is not telling us what our pre-existing concept of justification amounts to, nor what is essential to the referent of that concept. Rather, in a revisionary spirit, it is proposing a new notion of justification. Actually, the divide between the first and the last of these alternatives is not all that sharp, for one might start with an analysis of our ordinary notion and go on to propose modifications in an effort to tidy the notion up a bit here and there. As the changes proposed get bigger and bigger, this sort of "explication" gradually shades into pure stipulation. So long as the changes an explication urges in a pre-existing concept are motivated by considerations of simplicity and don't result in any radical departures from the ordinary concept, I'll count them as a kind of conceptual analysis. I think a good case can be made that Goodman took himself to be providing just such a conservative explication. But again, since it is a view rather than an author that I hope to refute, I will simply stipulate that the conceptual analysis or conservative explication interpretation is the one to be adopted here.

3 Does the Reflective Equilibrium Account Capture our Notion of Justification?

Goodman, as I propose to read him, offers us an account of what our concept of justified inference comes to. How can we determine whether his analysis is correct? One obvious strategy is to ask just what systems of inferential rules result from the process of mutual adjustment that Goodman advocates. If the inferential systems generated by the reflective equilibrium process strike us as systems that a rational person ought to invoke, this will count in favor of Goodman's analysis. If, on the other hand, the reflective equilibrium process generates what we take to be irrational or unjustified inferential rules or practices, this will cast doubt on Goodman's claim to have captured our concept of justification. Since we are viewing conceptual explication as a kind of analysis, we should

not insist that Goodman's account coincide perfectly with our intuitive judgments. But if there are lots of cases in which Goodman's account entails that a system of inferential rules is justified and intuition decrees that it is not, this is a symptom that the analysis is in serious trouble.

In an earlier paper, Nisbett and I exploited the strategy just described to argue that the reflective equilibrium account does not capture anything much like our ordinary notion of justification.¹⁰ On the basis of both controlled studies and anecdotal evidence, we argued that patently unacceptable rules of inference would pass the reflective equilibrium test for many people. For example, it appears likely that many people infer in accordance with some version of the gambler's fallacy when dealing with games of chance. These people infer that the likelihood of throwing a seven in a game of craps increases each time a non-seven is thrown. What is more, there is every reason to think that the principle underlying their inference is in reflective equilibrium for them. When the principle is articulated and the subjects have had a chance to reflect upon it and upon their own inferential practice, they accept both. Indeed, one can even find some nineteenth century logic texts in which versions of the gambler's fallacy are explicitly endorsed. (In a delightful irony, one of these books was written by a man who held the same chair Goodman held when he wrote *Fact, Fiction and Forecast*.¹¹) It can also be shown that many people systematically ignore the importance of base rates in their probabilistic reasoning, that many find the principle of regression to the mean to be highly counter-intuitive, that many judge the probability of certain sequences of events to be higher than the probability of components in the sequence, etc.¹² In each of these cases, and in many more that might be cited, it is very likely that, for some people at least, the principles that capture their inferential practice would pass the reflective equilibrium test. If this is right, it indicates there is something very wrong with the Goodmanian analysis of justification. For on that analysis, to be justified *is* to pass the reflective equilibrium test. But few of us are prepared to say that if the gambler's fallacy is in reflective equilibrium for a person, then his inferences that accord with that principle are justified.

Of course, each example of the infelicitous inferential principle that allegedly would pass the reflective equilibrium test is open to challenge. Whether or not the dubious principles that appear

to guide many people's inferential practice would stand up to the reflective scrutiny Goodman's test demands is an empirical question. And for any given rule, a Goodmanian might protest that the empirical case has just not been made adequately. I am inclined to think that the Goodmanian who builds his defenses here is bound to be routed by a growing onslaught of empirical findings. But the issue need not turn on whether this empirical hunch is correct. For even the *possibility* that the facts will turn out as I suspect they will poses a serious problem for the Goodmanian story. It is surely not an a priori fact that strange inferential principles will always fail the reflective equilibrium test for all subjects. And if it is granted, as surely it must be, that the gambler's fallacy (or any of the other inferential oddities that have attracted the attention of psychologists in recent years) could possibly pass the reflective equilibrium test for some group of subjects, this is enough to cast doubt on the view that reflective equilibrium is constitutive of justification as that notion is ordinarily used. For surely we are not at all inclined to say that a person is justified in using any inferential principle – no matter how bizarre it may be – simply because it accords with his reflective inferential practice.

Faced with this argument the friends of reflective equilibrium may offer a variety of responses. The one I have the hardest time understanding is simply to dig in one's heels and insist that if the gambler's fallacy (or some other curious principle) is in reflective equilibrium for a given person or group, then that principle is indeed justified for them. Although I have heard people advocate this line in conversation, I know of no one who has been bold enough to urge the view in print. Since no one else seems willing to take the view seriously, I won't either.

A very different sort of response is to urge that the notion of reflective equilibrium is itself in need of patching – that some bells and whistles must be added to the justificatory process Goodman describes. One idea along these lines is to shift from narrow Goodmanian reflective equilibrium to some analog of Rawls's "wide reflective equilibrium."¹³ Roughly, the idea here is to broaden the scope of the judgments and convictions that are to be brought into coherence with one another. Instead of attending only to our assessments of inferential principles, wide reflective equilibrium also requires that our system of inferential rules is to cohere with our semantic, epistemological,

metaphysical, or psychological views. Just how various philosophical or psychological convictions are supposed to constrain a person's inferential principles and practice has not been spelled out in much detail, though Norman Daniels, whose papers on wide reflective equilibrium are among the best around, gives us a hint when he suggests, by way of example, that Dummett's views on logic are constrained by his semantic views.¹⁴ It would also be plausible to suppose that the classical intuitionists in logic rejected certain inferential principles on epistemological grounds.

A rather different way of attempting to preserve a reflective equilibrium account of justification is to restrict the class of people whose reflective equilibrium is to count in assessing the justification of inferential principles. For example, Nisbett and I proposed that in saying an inferential principle is justified, what we are saying is that it would pass the (narrow) reflective equilibrium test for those people whom we regard as experts in the relevant inferential domain.¹⁵

A dubious virtue of both the wide reflective equilibrium and the expert reflective equilibrium accounts is that they make clear-cut counter-examples harder to generate. That is, they make it harder to produce actual examples of inferential rules which the analysis counts as justified and intuition does not. In the case of wide reflective equilibrium, counter-examples are hard to come by just because it is so hard to show that anything is in wide reflective equilibrium for anyone. ("Would she continue to accept that rule if she thought through her epistemological and metaphysical views and came to some stable equilibrium view?" Well, God knows.) In the case of the expert reflective equilibrium account, the dubious but reflectively self-endorsed inferential practice of the experimental subject or the Las Vegas sucker just don't count as counter-examples, since these people don't count as experts.

But though clear-cut cases involving actual people may be harder to find, each of these elaborations of the reflective equilibrium story falls victim to the argument from possible cases offered earlier. Consider wide reflective equilibrium first. No matter how the details of the wide reflective equilibrium test are spelled out, it is surely not going to turn out to be impossible for a person to reach wide reflective equilibrium on a set of principles and convictions that includes some quite daffy inferential rule. Indeed, one suspects that by allowing people's philosophical convictions to

play a role in filtering their inferential principles, one is inviting such daffy principles, since many people are deeply attached to outlandish philosophical views. The expert reflective equilibrium move fares no better. For unless experts are picked out in a question-begging way (e.g. those people whose inferential practices are in fact justified) it seems entirely possible for the expert community, under the influence of ideology, recreational chemistry, or evil demons, to end up endorsing some quite nutty set of rules.¹⁶

4 A "Neo-Goodmanian" Project

At this point, if the friend of reflective equilibrium is as impressed by these arguments as I think he should be, he might head off to his study to work on some further variations on the reflective equilibrium theme that will do better at capturing our concept of justification. Despite a string of failures, he might be encouraged to pursue this project by a line of thought that runs something like the following. I'll call it the *neo-Goodmanian* line.

It can hardly be denied that we do *something* to assess whether or not an inferential practice is justified. Our decisions on these matters are certainly not made at random. Moreover, if there is some established procedure that we invoke in assessing justification, then it must surely be possible to describe this procedure. When we have succeeded at this we will have an account of what it is for an inferential practice to be justified. For, as Goodman has urged, to be justified just *is* to pass the tests we invoke in assessing an inferential practice. Our procedures for assessing an inferential practice are constitutive of justification. Granted, neither Goodman's narrow reflective equilibrium story nor the more elaborate stories told by others has succeeded in capturing the procedure we actually use in assessing justification. But that just shows we must work harder. The rewards promise to repay our efforts, since once we have succeeded in describing our assessment procedure, we will have taken a giant step forward in epistemology. We will have explained what it is for a cognitive process to be justified. In so doing we will have at least begun to resolve the problem posed by cognitive diversity. For once we have a clear specification of what justification amounts to, we can go on to ask whether our

own cognitive processes are justified or whether, perhaps, those of some other culture come closer to the mark.

There is no doubt that this neo-Goodmanian line can be very appealing. I was myself under its sway for some years. However, I am now persuaded that the research program it proposes for epistemology is a thoroughly wrong-headed one. In the pages that follow I will try to say why. My case against the neo-Goodmanian project divides into two parts. First I shall raise some objections that are targeted more or less specifically on the details of the neo-Goodmanian program. Central to each of these objections is the fact that the neo-Goodmanian is helping himself to a healthy serving of empirical assumptions about the conceptual structures underlying our commonsense judgments of cognitive assessment, and each of these assumptions stands in some serious risk of turning out to be false. If one or more of them is false, then the project loses much of its initial attractiveness. In the following section I will set out a brief catalog of these dubious assumptions. The second part of my critique is much more general and I'll be after much bigger game. What I propose to argue is that neither the neo-Goodmanian program nor any alternative program that proposes to analyze or explicate our pre-systematic notion of justification will be of any help at all in resolving the problem posed by cognitive diversity. But here I am getting ahead of myself. Let me get back to the neo-Goodmanian and his dubious empirical presuppositions.

5 Some Questionable Presuppositions of the Neo-Goodmanian Project

Let me begin with a fairly obvious point. The neo-Goodmanian, as I have portrayed him, retains his allegiance to the idea of reflective equilibrium. We last saw him heading back to his study to seek a more adequate elaboration of this notion. But nothing the neo-Goodmanian has said encourages us to expect that reflective equilibrium or anything much like it plays a role in our procedure for assessing the justification of a cognitive process. So even if it is granted that we have good reason to work hard at characterizing our justification-assessing procedure, we may find that the notion of reflective equilibrium is simply a non-starter. Confronted with this objection, I think the only

move open to the neo-Goodmanian is to grant the point and concede that in trying to patch the notion of reflective equilibrium he is simply playing a hunch. Perhaps it will turn out that something like reflective equilibrium plays a central role in our assessments of justification. But until we have an accurate characterization of the assessment process there can be no guarantees.

Two further assumptions of the neo-Goodmanian program are that we ordinarily invoke only *one* notion of justification for inferential processes, and that this is a *coherent* notion for which a set of necessary and sufficient conditions can be given. But once again these are not matters that can be known in advance. It might be that different people mean different things when they call a cognitive process “justified” because there are different notions of justification in circulation. These different meanings might cluster around a central core. But then again they might not. There are lots of normatively loaded terms that seem to be used in very different ways by different individuals or groups in society. I would not be at all surprised to learn that what I mean by terms like “morally right” and “freedom” is very different from what the followers of the Revd Falwell or admirers of Col. Khadafi mean. And I wouldn’t be much more surprised if terms of epistemic evaluation turned out to manifest similar interpersonal ambiguities.

Even discounting the possibility of systematic interpersonal differences, it might be that in assessing the justification of a cognitive process we use different procedures on different occasions, and that these procedures have different outcomes. Perhaps, for example, our intuitive notion of justification is tied to a number of prototypical exemplars, and that in deciding new cases we focus in some context-sensitive way on one or another of these exemplars, making our decision about justification on the basis of how similar the case at hand is to the exemplar on which we are focusing. This is hardly a fanciful idea, since recent work on the psychological mechanisms underlying categorization suggests that in *lots* of cases our judgment works in just this way.¹⁷ If it turns out that our judgments about the justification of cognitive processes are prototype or exemplar based, then it will be a mistake to look for a property or characteristic that all justified cognitive processes have. It will not be the case that there is any single test passed by all the cognitive processes we judge to be justified. I am partial to a reading of the later Wittgenstein on which this is just what he would urge

about our commonsense notion of justification, and I am inclined to suspect that this Wittgensteinian story is right. But I don’t pretend to have enough evidence to make a convincing case. For present purposes it will have to suffice to note that this *might* be how our commonsense concept of justification works. If it is, then the neo-Goodmanian program is in for some rough sledding.

A final difficulty with the neo-Goodmanian program is that it assumes, without any evidence, that the test or procedure we use for assessing the justification of cognitive processes exhausts our concept of inferential justification, and thus that we will have characterized the concept when we have described the test. But this is hardly a claim that can be assumed without argument. It might be the case that our procrustean concept of justification is an amalgam composed in part of folk epistemological theory specifying certain properties or characteristics that are essential to justification, and in part of a test or cluster of tests that folk wisdom holds to be indicative of those properties. Moreover, the tests proposed might not always (or ever) be reliable indicators of the properties.¹⁸ I don’t have any compelling reason to believe that our commonsense notion of justification will turn out like this. But I wouldn’t be much surprised. Though our understanding of the mechanisms underlying commonsense concepts and judgments is still *very* primitive, as I read the literature it points to two important morals. First, the mental representation of concepts is likely to turn out to be a very messy business. Second, it is no easy job to separate commonsense concepts from the folk theories in which they are enmeshed. All of this bodes ill for the neo-Goodmanian who hopes that the analysis or explication of our concept of justification will yield some relatively straightforward elaboration of the reflective equilibrium test.

6 Against Analytic Epistemology

The problems posed in the previous section shared a pair of properties. They all turned on empirical assumptions about the nature of our ordinary concept of justification, and they were all targeted fairly specifically at the neo-Goodmanian project.¹⁹ In the current section I want to set out a very different sort of argument, an argument which if successful will undermine not only reflective equilibrium theories but also the whole family of epistemological theories to which they belong.

To give some idea of the range of theories that are in the intended scope of my critique, it will be helpful to sketch a bit of the framework for epistemological theorizing suggested by Alvin Goldman in his recent book, *Epistemology and Cognition*.²⁰ Goldman notes that one of the major projects of both classical and contemporary epistemology has been to develop a theory of epistemic justification. The ultimate job of such a theory is to say which cognitive states are epistemically justified and which are not. Thus, a fundamental step in constructing a theory of justification will be to articulate a system of rules evaluating the justificatory status of beliefs and other cognitive states. These rules (Goldman calls them *justificational rules* or *J-rules*) will specify permissible ways in which a cognitive agent may go about the business of forming or updating his cognitive states. They “permit or prohibit beliefs, directly or indirectly, as a function of some states, relations, or processes of the cognizer.”²¹

Of course, different theorists may have different views on which beliefs are justified or which cognitive processes yield justified beliefs, and thus they may urge different and incompatible sets of J-rules. It may be that there is more than one right system of justificational rules, but it is surely not the case that all systems are correct. So in order to decide whether a proposed system of J-rules is right, we must appeal to a higher criterion which Goldman calls a “criterion of rightness.” This criterion will specify a “set of conditions that are necessary and sufficient for a set of J-rules to be right.”²²

But now the theoretical disputes emerge at a higher level, for different theorists have suggested very different criteria of rightness. Indeed, as Goldman notes, an illuminating taxonomy of epistemological theories can be generated by classifying theories or theorists on the basis of the sort of criterion of rightness they endorse. Coherence theories, for example, take the rightness of a system of J-rules to turn on whether conformity with the rules would lead to a coherent set of beliefs. Truth linked or reliability theories take the rightness of a set of J-rules to turn in one way or another on the truth of the set of beliefs that would result from conformity with the rules. Reflective equilibrium theories judge J-rules by how well they do on their favored version of the reflective equilibrium test. And so on. How are we to go about deciding among these various criteria of rightness? Or, to ask an even more basic ques-

tion, just what does the correctness of a criterion of rightness come to; what makes a criterion right or wrong? On this point Goldman is not as explicit as one might wish. However, much of what he says suggests that, on his view, *conceptual analysis* or *conceptual explication* is the proper way to decide among competing criteria of rightness. The correct criterion of rightness is the one that comports with the conception of justifiedness that is “embraced by everyday thought or language.”²³ To test a criterion we explore the judgments it would entail about specific cases, and we test these judgments against our “pretheoretic intuition.” “A criterion is supported to the extent that implied judgments accord with such intuitions, and weakened to the extent that they do not.”²⁴ Goldman is careful to note that there may be a certain amount of vagueness in our commonsense notion of justifiedness, and thus there may be no unique best criterion of rightness. But despite the vagueness, “there seems to be a common core idea of justifiedness” embedded in everyday thought and language, and it is this common core idea that Goldman tells us he is trying to capture in his own epistemological theorizing.²⁵

The view I am attributing to Goldman on what it is for a criterion of rightness to itself be right is hardly an idiosyncratic or unfamiliar one. We saw earlier that a very natural reading of Goodman would have him offering the reflective equilibrium story as an explication or conceptual analysis of the ordinary notion of justification. And many other philosophers have explicitly or implicitly adopted much the same view. I propose to use the term *analytic epistemology* to denote any epistemological project that takes the choice between competing justificational rules or competing criteria of rightness to turn on conceptual or linguistic analysis. There can be little doubt that a very substantial fraction of the epistemological writing published in English in the last quarter of a century has been analytic epistemology.²⁶ However, it is my contention that if an analytic epistemological theory is taken to be part of the serious normative inquiry whose goal is to tell people which cognitive processes are good ones, or which ones they should use, then for most people it will prove to be an irrelevant failure.

I think the most intuitive way to see this point is to begin by recalling how the specter of culturally based cognitive diversity lends a certain urgency to the question of which cognitive processes we should use. If patterns of inference are acquired

from the surrounding culture, much as language or fashions or manners are, and if we can learn to use cognitive processes quite different from the ones we have inherited from our culture, then the question of whether our culturally inherited cognitive processes are good ones is of more than theoretical interest. If we *can* go about the business of cognition differently, and if others actually *do*, it is natural to ask whether there is any reason why we should continue to do it our way. Even if we cannot change our cognitive processes once we've acquired them, it is natural to wonder whether those processes are good ones. Moreover, for many people the absence of a convincing affirmative answer can be seriously disquieting. For if we cannot say why our cognitive processes are any better than those prevailing elsewhere, it suggests that it is ultimately no more than an historical accident that we use the cognitive processes we do, or that we hold the beliefs that those processes generate, just as it is an historical accident that we speak English rather than Spanish and wear trousers rather than togas.

Consider now how the analytic epistemologist would address the problem that cognitive diversity presents. To determine whether our cognitive processes are good ones, he would urge, we must first *analyze* our concept of justification (or perhaps some other commonsense epistemic notion like rationality). If our commonsense epistemic notion is not too vague or ambiguous, the analysis will give us a criterion of rightness for J-rules (or perhaps a cluster of closely related criteria). Our next step is to investigate which sets of J-rules fit the criterion. Having made some progress there, we can take a look at our own cognitive processes and ask whether they do in fact accord with some right set of J-rules. If they do, we have found a reason to continue using those processes; we have shown that they are good ones because the beliefs they lead to are justified. If it turns out that our cognitive processes don't accord with a right set of J-rules, we can try to discover some alternative processes that do a better job, and set about training ourselves to use them.

It is my contention that something has gone very wrong here. For the analytic epistemologist's effort is designed to determine whether our cognitive states and processes accord with our commonsense notion of justification (or some other commonsense concept of epistemic evaluation). Yet surely the evaluative epistemic concepts embedded in everyday thought and language are

every bit as likely as the cognitive processes they evaluate to be culturally acquired and to vary from culture to culture.²⁷ Moreover, the analytic epistemologist offers us no reason whatever to think that the notions of evaluation prevailing in our own language and culture are any better than the alternative evaluative notions that might or do prevail in other cultures. But in the absence of any reason to think that the locally prevailing notions of epistemic evaluation are superior to the alternatives, why should we care one whit whether the cognitive processes we use are sanctioned by those evaluative concepts? How can the fact that our cognitive processes are approved by the evaluative notions embraced in our culture alleviate the worry that our cognitive processes are no better than those of exotic folk, if we have no reason to believe that our evaluative notions are any better than alternative evaluative notions?

To put the point a bit more vividly, imagine that we have located some exotic culture that does in fact exploit cognitive processes very different from our own, and that the notions of epistemic evaluation embedded in their language also differ from ours. Suppose further that the cognitive processes prevailing in that culture accord quite well with *their* evaluative notions, while the cognitive processes prevailing in our culture accord quite well with *ours*. Would any of this be of any help at all in deciding which cognitive processes we should use? Without some reason to think that one set of evaluative notions was preferable to the other, it seems clear that it would be of no help at all.

In the philosophical literature there is a tradition, perhaps traceable to Wittgenstein, that would reject the suggestion that our evaluative notions should themselves be evaluated. Justifications, this tradition insists, must come to an end. And once we have shown that our practice accords with our evaluative concepts, there is nothing more to show. Our language game (or form of life) does not provide us with any way to go about evaluating our evaluative notions. There is no logical space in which questions like "should we hold justified beliefs?" or "should we invoke rational cognitive processes?" can be asked seriously. If a person did not recognize that the answers to these questions had to be affirmative, it would simply indicate that he did not understand the logical grammar of words like "should" and "justified" and "rational."

I am inclined to think that there is at least a kernel of truth in this "Wittgensteinian" stand.

Justifications do ultimately come to an end. However, it is, I think, a disastrous mistake to think that they come to an end *here*. For there are *lots* of values that are both widely shared and directly relevant to our cognitive lives, though they are quite distinct from the “epistemic values” that lie behind our ordinary use of terms like “justified” and “rational.” It is against the background of these non-epistemic values that our socially shared system of epistemic evaluation can itself be evaluated. Thus, for example, many people attach high value to cognitive states that foster happiness (their own or everyone’s), and many people value cognitive states that afford them the power to predict and control nature. Some people share Mother Nature’s concern that our cognitive lives should foster reproductive success. And, on a rather different dimension, many people care deeply that their beliefs be true.²⁸ Each of these values, along with many others that might be mentioned, affords a perspective from which epistemic values like justification and rationality can be evaluated. We can ask whether the cognitive states and processes endorsed by our notions of epistemic value foster happiness, or power, or accurate prediction, or reproductive success, or truth. More interestingly, we can ask whether the cognitive states and processes we actually have or use foster happiness, power, or the rest. And if they do not, we can explore alternatives that may do a better job, though there is of course no guarantee that all of these values can be maximized together.²⁹

At this point, it might be protested that the values I am proposing to use in evaluating our socially shared notions of epistemic evaluation are themselves lacking any deeper justification. If someone can accept *these* as ultimate values, why couldn’t someone do the same for justification or rationality? My reply is that of course someone could, but this is no objection to the view I am urging. There are many things that people might and do find ultimately or intrinsically valuable. Some of these values may be rooted more or less directly in our biological nature, and these we can expect to be widely shared. Other values, including intrinsic, life-shaping values, may be socially transmitted, and vary from society to society. Still others may be quite idiosyncratic. It is entirely possible for someone in our society to attach enormous value to having justified beliefs or to using rational inferential strategies – that is, to having beliefs or inferential processes that fall within the extension of “justified” or “rational” as they are

used in our language. Similarly, it is entirely possible for someone in another society to attach enormous value to having cognitive states that fall within the extension of the terms of cognitive evaluation current in that society. In each case the evaluation may be either instrumental or intrinsic. A person in our culture may value the states and processes that fall within the extension of “rational” or “justified” because he thinks they are likely to be true, to lead to happiness, etc., or he may value them for no further reason at all. And a person in another culture may have either sort of attitude in valuing what falls within the extension of his language’s terms of cognitive evaluation. Where the value attached is instrumental, there is plenty of room for productive inquiry and dialogue. We can try to find out whether rational or justified cognitive processes do lead to happiness or power or truth, and if they do we can try to understand why. But where the value accorded to one or another epistemic virtue is intrinsic, there is little room for debate. If you value rationality for its own sake, and the native of another culture values some rather different cognitive characteristic (“shmashinality” as Hilary Putnam might put it) for *its* own sake, there is not much you can say to each other. Moreover, there is not much I can say to either of you, since on my view the fact that a cognitive process is sanctioned by the venerable standards embedded in our language of epistemic evaluation, or theirs, is of no more interest than the fact that it is sanctioned by the venerable standards of a religious tradition or an ancient text – unless, of course, it can be shown that those standards correlate with something more generally valued.³⁰ But I do not pretend to have any arguments that will move the true epistemic xenophobe. If a person really does attach deep intrinsic value to the epistemic virtues favored by folk epistemology, then dialogue has come to an end.

Finally, let me say how all of this relates to analytic epistemology. The analytic epistemologist proposes to arbitrate between competing criteria of rightness by seeing which one accords best with the evaluative notions “embraced by everyday thought and language.” However, it is my contention that this project is of no help whatever in confronting the problem of cognitive diversity unless one is an epistemic xenophobe. The program of analytic epistemology views conceptual analysis or explication as a stopping place in disputes about how we should go about the business of cognition. When we know that a certain cognit-

ive process falls within the extension of our ordinary terms of epistemic evaluation – whatever the analysis of those terms may turn out to be – we know all that can be known that is relevant to the questions of how we should go about the business of reasoning. But as I see it, the only people who should take this information to be at all relevant to the question are the profoundly conservative people who find intrinsic value in having their cognit-

ive processes sanctioned by culturally inherited standards, whatever those standards may be. Many of us care very much whether our cognitive processes lead to beliefs that are true, or give us power over nature, or lead to happiness. But only those with a deep and free-floating conservatism in matters epistemic will care whether their cognitive processes are sanctioned by the evaluative standards that happen to be woven into our language.

Notes

- 1 See Davidson (1982); Stich (1979); Routley (1981); Stich (1983), pp. 89–106; Stich (1984).
- 2 Nor are these the only alternatives. There are lots of characteristics which are innate (not part of our cultural heritage) though they differ substantially from one group to another. Sex, hair color, and blood type are three obvious examples.
- 3 See Cole and Scribner (1974); Cole and Means (1981).
- 4 Actually, the issue is not so straightforward if we compare languages at very different stages of development, or languages involving different theoretical assumptions. It is only when the choice is between languages that are more or less inter-translatable with our own that we are inclined to judge that one is as good as another. Thanks to Paul Churchland for reminding me of this point.
- 5 Rawls (1971), pp. 20ff.
- 6 Goodman (1965), pp. 66–7; emphasis is Goodman's.
- 7 Cherniak (1986, ch. 4); Harman (1986, ch. 2); Goldman (1986, sec. 5.1).
- 8 L. J. Cohen (1981) seems to read Goodman this way since he exploits Goodman's notion reflective equilibrium in giving an account of good reasoning.
- 9 Well, I will argue it a little. Note first that according to Goodman the only justification needed for either rules or inferences "lies in" the agreement achieved by the reflective equilibrium process. This talk of justification *lying in* the agreement strongly suggests the constitutive reading. Moreover, on the non-constitutive reading, Goodman's doctrine would be an oddly incomplete one. It would present us with a test for justification without telling us why it was a test or giving us any account of what it is that is being tested for. On the constitutive reading, by contrast, no such problem arises. We have in one tidy package both an analysis of the notion of justification and an unproblematic explanation of the relation between justification and the process Goodman describes.
- 10 Stich and Nisbett (1980).
- 11 The writer was Henry Coppée (1874). Here is a brief quote:

Thus, in throwing dice, we cannot be sure that any single face or combination of faces will appear; but if, in very many throws, some particular face has not appeared, the chances of its coming up are stronger and stronger, until they approach very near to certainty. It must come; and as each throw is made and it fails to appear, the certainty of its coming draws nearer and nearer. (p. 162)
- 12 For an excellent survey of the literature in this area see Nisbett and Ross (1980); a number of important studies are collected in Kahneman, Slovic, and Tversky (1982).
- 13 Rawls (1974).
- 14 Daniels (1979, 1980a, 1980b).
- 15 Stich and Nisbett (1980).
- 16 As Conee and Feldman (1983) point out, the situation is actually a bit worse for the version of the expert reflective equilibrium analysis that Nisbett and I offered. On that account, different groups may recognize different people as experts. And it is surely at least possible for a group of people to accept as an expert some guru who is as bonkers as he is charismatic. But we certainly don't want to say that the followers of such a guru would be rational to invoke whatever wild inferential principle might be in reflective equilibrium for their leader.
- 17 For a good review of the literature, see Smith and Medin (1981).
- 18 For some insightful observations on the potential complexity of commonsense concepts and the ways in which intuitive tests can fail to capture the extension of concepts, see Rey (1983).
- 19 Actually, the last three of my four objections might, with a bit of reworking, be generalized so as to apply to all of analytic epistemology, as it is defined below. But I don't propose to pursue them since, as we shall

see, analytic epistemology has more pressing problems.

20 Goldman (1986).

21 Ibid., p. 60. For the reader who wants a more hands-on feel for Goldman's notion of a J-rule, the quote continues as follows:

For example, J-rules might permit a cognizer to form a given belief because of some appropriate antecedent or current state. Thus, someone being "appeared to" in a certain way at t might be permitted to believe p at t . But someone else not in such a state would not be so permitted. Alternatively, the rules might focus on mental operations. Thus, if S 's believing p at t is the result of a certain operation, or sequence of operations, then his belief is justified if the system of J-rules permits that operation or sequence of operations.

22 Ibid., p. 64.

23 Ibid., p. 58.

24 Ibid., p. 66.

25 Ibid., pp. 58–9.

26 For an extended review of part of this literature see Shope (1983). As Shope notes, relatively few of the philosophers who have tried their hands at constructing an "analysis" of knowledge (or of some other epistemic notion) have been explicit about their objectives (see pp. 34–44). However, absent indications to the contrary, I am inclined to think that if a philosophical project proceeds by offering definitions or "truth conditions," and testing them against our intuitions about real or imaginary cases, then the project should be viewed as an attempt at conceptual analysis or explication. Unless one has some pretty strange views about intuitions, it is hard to see what we could hope to gain from capturing them apart from some insight into the concepts that underlie them.

27 Evidence on this point, like evidence about cross-cultural differences in cognitive processes, is hard to come by and hard to interpret. But there are some intriguing hints in the literature. Hallen and Sodipo (1986) studied the terms of epistemic evaluation exploited by the Yoruba, a West African people. It is their contention that the Yoruba do not have a distinction corresponding to our distinction between knowledge and (mere) true belief. They do, however, divide beliefs into two other categories: those for which a person has immediate, eyewitness evidence, and those for which he does not. In the standard Yoruba–English dictionaries, the Yoruba term for the former sort of belief, "mo," is translated as "knowledge" while the term for the latter sort, "gbagbo," is translated as "belief." However, Hallen and Sodipo argue that these translations are mistaken, since "mo" has a much narrower extension than "knowledge." Most of what we would

classify as scientific knowledge, for example, would not count as "mo" for the Yoruba, because it is based on inference and secondhand report. Since the Yoruba do not draw the distinction between knowledge and (mere) true belief, they have no use for our notion of epistemic justification, which earns its keep in helping to draw that distinction. Instead, the Yoruba presumably have another notion which they exploit in distinguishing "mo" from "gbagbo." Hallen and Sodipo do not indicate whether the Yoruba have a single word for this notion, but if they do, it would be a mistake to translate the word as "(epistemic) justification." Clearly, if Hallen and Sodipo are right, the Yoruba categories of epistemic evaluation are significantly different from our own.

28 I should note, in passing, that I think it is a mistake to include truth on the list of intrinsically valuable features of one's cognitive life. But that is a topic for another paper, (see Stich, in preparation) and I will ignore the point here.

29 The point I am making here is really just a generalization of a point made long ago by Salmon (1957), Skyrms (1975) and a number of other authors. Strawson (1952) argued that the rationality or reasonableness of inductive reasoning was easy to demonstrate, since being supported by inductive inference is part of what we *mean* when we say that an empirical belief is *reasonable*. To which Salmon replied that if Strawson is right about the meaning of "reasonable" it is not at all clear why anyone should *want* to be reasonable. What most of us do care about, Salmon notes, is that our inferential methods be those that are "best suited to the attainment of our ends" (p. 41). "If we regard beliefs as reasonable simply because they are arrived at inductively and we hold that reasonable beliefs are valuable for their own sake, it appears that we have elevated inductive method to the place of an intrinsic good" (p. 42). The analytic epistemologist elevates being within the extension of our ordinary terms of epistemic evaluation to the place of an intrinsic good. In so doing, the analytic epistemologist embraces a system of value that few of us are willing to share.

30 Let me try to head off a possible misunderstanding. Some analytic epistemologists claim that our ordinary notions of epistemic evaluation are conceptually linked to truth. On Goldman's account, for example, the rightness of a set of J-rules is a function of how well the processes sanctioned by those rules do at producing truths. If this is right, then a person who attached intrinsic value to having true beliefs would, of course, have reason to be interested in whether his cognitive states and processes were sanctioned by the standards embedded in our language. But here it is the appeal to truth that is doing the work, not the appeal to traditional standards. For if Goldman is wrong in his conceptual analysis and "(epistemic) justification" is not conceptually tied to truth, the

person who values truth will stay just as interested in whether his cognitive processes reliably lead to truth, though he may have no interest whatever in

how traditional notions of epistemic evaluation judge his cognitive processes. Thanks to Steven Luper-Foy for the query that prompted this note.

References

- Cherniak, C., 1986. *Minimal Rationality* (Cambridge, MA: MIT Press).
- Cohen, L. J., 1981, "Can Human Irrationality Be Experimentally Demonstrated?," *Behavioral and Brain Sciences* 4, pp. 317–70.
- Cole, M. and S. Scribner, 1974. *Culture and Thought* (New York: John Wiley).
- Cole, M. and B. Means: 1981, *Comparative Studies of How People Think* (Cambridge, MA: Harvard University Press).
- Conce, E. and R. Feldman, 1983, "Stich and Nisbett on Justifying Inference Rules," *Philosophy of Science* 50, 326–31.
- Daniels, N., 1979, "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *Journal of Philosophy* 76, pp. 256–82.
- , 1980a. "Reflective Equilibrium and Archimedean Points," *Canadian Journal of Philosophy* 10, pp. 83–103.
- , 1980b. "On Some Methods of Ethics and Linguistics," *Philosophical Studies*, 37, pp. 21–36.
- Davidson, D., 1982. "Rational Animals," *Dialectica* 36, pp. 317–27.
- Goldman, A., 1986, *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- Goodman, N., 1965. *Fact, Fiction and Forecast* (Indianapolis: Bobbs-Merrill).
- Hallen, B. and J. O. Sodipo, 1986. *Knowledge, Belief and Witchcraft* (London: Ethnographica).
- Harman, G., 1986. *Change of View* (Cambridge, MA: MIT Press).
- Kahneman, D., P. Slovic, and A. Tversky, 1982. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge, MA: Cambridge University Press).
- Nisbett, R. and L. Ross, 1980. *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, NJ: Prentice-Hall).
- Rawls, J., 1971. *A Theory of Justice* (Cambridge, MA: Harvard University Press).
- , 1974. "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association* 48, pp. 4–22.
- Rey, G., 1983. "Concepts and Stereotypes," *Cognition* 15, pp. 237–62.
- Routley, R., 1981. "Alleged Problems Attributing Beliefs, and Intentionality, to Animals," *Inquiry* 24, pp. 385–417.
- Salmon, W., 1957. "Should We Attempt to Justify Induction?" *Philosophical Studies* 8, pp. 33–48.
- Shope, R., 1983. *The Analysis of Knowing* (Princeton: Princeton University Press).
- Skyrms, B., 1975. *Choice and Chance* (Belmont, CA: Wadsworth).
- Smith, E. and D. Medin, 1981. *Concepts and Categories* (Cambridge, MA: Harvard University Press).
- Stich, S., 1979. "Do Animals Have Beliefs?," *Australasian Journal of Philosophy* 57, 15–28.
- , 1983. *From Folk Psychology to Cognitive Science* (Cambridge, MA: MIT Press).
- , 1984. "Relativism, Rationality and the Limits of Intentional Description," *Pacific Philosophical Quarterly* 65, pp. 211–235.
- , in preparation, "Do You Really Care Whether Your Beliefs Are True?"
- Stich, S. and R. Nisbett, 1980. "Justification and the Psychology of Human Reasoning," *Philosophy of Science* 47, pp. 188–202.
- Strawson, P., 1952. *Introduction to Logical Theory* (New York: John Wiley).

Index

- a posteriori* knowledge, 240, 254
a priori knowledge, 117, 131, 240, 254, 262, 265, concept possession, 218; epistemology and, 218, 404, 408, 573–4, 578, externalism, 391, intuition and evidence, 262, necessity and analyticity, 121–2, 218–9
- absolute terms, 46–7, 506, 517–8
- Academic skepticism, 114
- acceptance rules, 67, 111–3
- access, direct, 193, 197–8
- accidental: and knowledge, 60–1, 184, 278
- Ackerman, Diana, 223n
- Ackermann, Robert, 271n
- adequacy conditions justification, 557–61; perception, 329
- Agrippa's Trilemma, 281
- Alston, William P., 57, 106, 169, 222n, 235n, 240, 248–9, 251, 253–4, 258n, 259n, 260n, 272n, 311n, 337, 352n, 373, 375–8, 380, 384n, 385–6n, 410–1, 445, 455n, 569n
- Amico, Robert, 240
- analytic/synthetic distinction, 157, 298, 542–4; *see also* Quine
- Anderson, John, 222n
- animal knowledge, 239–40
- Annas, Julia, 4
- Annis, David, 481, 555n
- Anscombe, G. E. M., 570n
- appearances, in epistemology, 113 5, 403, 414–5, 418–9
- Aristotle, 109, 130, 271n, 285n, 330, 457
- Armstrong, D. M., 83, 95n, 98n, 101n, 152n, 180n, 259n, 266–7, 271n, 272n, 278, 285n, 290, 309, 312n, 371n, 387, 388, 392, 398n, 400n, 402, 405, 515n
- Arnauld, 256n
- artificial intelligence (AI): and epistemology, 196
- assertability: contextualism, 495–6
- attitude, doxastic, 167, 170
- Audi, Robert, 106, 169, 290, 312n
- Augustine, St., 114, 119n
- Aune, Bruce, 142
- Austin, J. L., 98n, 125, 316, 448, 481, 545
- awareness, 128, 239, 472
- Axtell, Guy, 436
- Ayer, A. J., 58, 59n, 66n, 273n, 308, 311n, 423n, 527
- Bach, Kent, 223n
- Bacon, Francis, 542, 554n
- Baergen, Ralph, 376, 379–80, 384n, 386n
- Baker, Gordon, 569n
- Baltimore, David, 446, 455n
- Barnes, Jonathan, 4
- Barrett, R., 235n
- Bayes and Bayesianism, 155, 448, 512
- Beardsmore, R. E., 235n
- belief and beliefs: apt, 278–9, 283; basic, 36–41, 127, 135–6, 147–8, 226–7, 263–9, 302, 533, 547; and causation, 157, 331; degrees of, 67, 507–8; formation of inductive, 37; immediate, 265, 268, 401–5; justified true belief (JTB), 55, 60–2, 125, 262, and knowledge, 55, 60–2, 125, 262; memory, 117, 221, 229, 242–3, 247, method of, 56, 82–5, 93–5; native processes, 446–53; and rationality, 159–62, 289, 305–7, reliable, 123; self-justifying, 112–8, 124, 127–8; warranted, 65, 278–9, 283, 435–6, 445–6, 447–50, 453–5
- Bender, J., 106
- Bennett, Jonathan, 95n
- Bentham, Jeremy, 293, 295
- Berkeley, George, 34, 276
- Bernard, E. E., 300n
- Bieri, Peter, 41n
- Bigelow, John, 515n
- Black, Max, 119n
- Blackburn, Simon, 421n, 422n, 423n, 554n
- Blanshard, Brand, 111, 115, 116, 118n, 119n
- Blish, James, 99n
- BonJour, Laurence, 106, 152n, 176–8, 179n, 227, 232, 235n, 239, 240, 440–1, 462, 466n, 555n
- Bosanquet, Bernard, 272n
- Bourdan, 245
- Bouwisma, O. K., 119n

- Bandom, Robert, 338–9, 432n
 Brandt, Richard B., 312n
 Brueckner, A., 332n
 Buck, Roger C., 52n
 Burks, A., 236n
 Burnyeat, M. F., 4, 23n, 191n, 240, 555n
- Campbell, Donald, 300, 300n
 Canary, Catherine M., 332n
 Cargite, James, 52n
 Carnap, Rudolf, 33–6, 39, 41n, 111, 113, 118n, 207, 223n, 272n, 294–7, 300n, 303, 304, 311n, 318, 322, 552, 554
 Carrier, L. S., 101n
 Carroll, Lewis, 222n
 Carruthers, P., 569n
 Cartesian circle, 187–8, 239, 242–3, 246–8, 276
 Cartesian internalism, 182–3
 Castañeda, H.–N., 223n, 273n, 450, 456n
 causal theory, 70–1, 79, 91, 390–1
 Cavell, Stanley, 98n, 318, 319
 certainty, 29–32, 245–6
 Chakrabharti, A., 422n
 charity: principle of, 160, 307
 Chellas, Brian, 516n
 Cherniak, Christopher, 581n
 Chisholm, Roderick, 57, 58, 59n, 60, 65n, 66n, 105, 106, 125–6, 128, 130–1, 132n, 133n, 172, 173, 201, 223n, 224n, 239, 241, 249–52, 253–4, 256n, 258n, 259n, 264, 269, 271n, 273n, 285n, 304, 308, 311n, 312n, 334n, 342, 348, 403, 412n, 445–6, 515n, 529n, 569n
 Chomsky, Noam, 135, 315
 Churchland, Patricia, 456n
 Churchland, Paul, 581n
 circularity, epistemological, 94–5, 185, 187–8, 232, 281–4, 290, 408–12, 557–61
 clairvoyance, 176–7, 201, 234, 281–3, 440
 Clark, Romane, 223n
 Clarke, Thompson, 4, 98n, 552–4, 555n
 Clay, Marjorie, 5, 285n, 515n, 554n, 555n
 Clifford, W. K., 179n, 272n
 closure, epistemic, 89–92, 326–8, 390, 479, 512–3
 Coady, C. A. J., 515n
 Code, Lorraine, 436
cognitio, scientia and, 240, 282
 cognition: and direct realism, 199; epistemic, 249–53; goals, 193–6, 446–55, 458–66, 572
 cognitive capacities, 170, 233
 cognitive diversity, 448, 571–2, 576, 578–81
 cognitive equipment, human, 149–50, 184, 186, 233
 cognitive virtues, 439–43, 468–9
 Cohen, Jonathan, 222n
 Cohen, L. J., 581n
 Cohen, Marc, 66n
 Cohen, Robert S., 52n, 191n
 Cohen, Stewart, 180n, 181n, 394, 400n, 479, 480–1, 499n, 515n
 coherence, 117, 154, 198, 231, 263, 277, 279–80, 282, 304, 307, 393
 coherent systems objection, alternative, 154, 255
 coherentism, 105, 134–6, 168, 226–7, 239, 249–52, 254–5; and external world, 338; and foundationalism, 134–8, 143–6, 239; objections to, 134–8, 143–6, 239, 338; response to epistemic regress problem, 105, 138–9; theories of justification, 105, 347–8, 578; theories of truth, 110–1, 154
 Cole, M., 581n
 commitments and entitlements view of knowledge, 426, 428–31
 communal activity: knowledge as, 425–6, 428–30
 computers *see* artificial intelligence concepts, 168, 202–5; epistemological theory of, 204–5, 213–5; logical theory of, 202–4
 conceptual analysis, epistemic: 55, 204–5, 213–5, 218; and psychology, 308–11, 407–8
 Conee, Earl, 167, 337, 445, 455n, 581n
 context: and meaning, 504, 517–8, 545–6; role in justification, 549–54
 contextualism, 483–4, 517–28, 533, 549–54; attributor, 483; explanation of, 483–4, 517–8; and externalism, 506; and foundationalism, 546–9; and justification, 517–8, 549; and knowledge, 503–14, 517–28, 549–54; and skepticism, 479–80, 482–6, 487–96, 518–9, 549–54; warranted assertability maneuvers (WAM), 495–6
 contingent/necessary distinction, 121–2, 218–9
 Convention T, 155
 conversational context, 506–7, 517
 Coppee, Henry, 581n
 Cormman, James, 132n, 273n, 398n, 569n
 Cottingham, J., 284n
 Craig, Edward, 57, 285n
 Creary, L. G., 223n
 criterion: problem of, 105, 242
 critical cognitivism, 239, 253–5
- Dancy, Jonathan, 235n, 466n
 Daniels, Norman, 575
 Darwin, Charles, 314
 Davidson, Donald, 106, 285n, 289, 307, 312n, 332n, 352n, 456n, 581n
de se, 210–1
 defeaters, 56, 63, 77, 200; reliability, 329–32
 definition, knowledge, 55, 58–9, 62–3, 79
 Delaney, C. F., 191n
 Dennett, Daniel, 421n, 451, 456n
 deontologism, epistemic, 167
 DeRose, Keith, 4, 479–80, 499n, 500n, 501n, 523
 Descartes, René, 3, 6–23, 26, 27, 33–5, 65, 98n, 99n, 109, 134, 135, 136, 148, 167, 182–3, 185–8, 191n, 234, 239–40, 242–51, 253–4, 256n, 257n, 258n, 260n, 267, 276–8, 280–3, 284n, 285n, 294, 301–3, 308, 314, 322, 326, 327, 334n, 460, 498n, 530–1, 544, 548, 554n, 555n; internalism, 182–3; Evil Deceiver, 234, 245–6; foundationalism, 134, 276–7; mind/body dichotomy, 544; rationalism, 276–7; skepticism, 3, 6–23; *see also* Cartesian circle, Cartesian internalism, *cognitio*
 Devonshire, Duke of, 12, 17, 19, 29
 Dewey, John, 107, 109, 115, 116, 118, 118n, 119n, 132, 348, 459–63, 463, 466n, 555n
 Diamond, Cora, 515n
 Dimmet, Ernest, 465, 467n
 dogmatism, 35, 47–8, 50–2, 189, 248–9; in intuitionism, 254–5; and skepticism, 42, 46–8, 426
 Doney, Willis, 242–3, 256n, 258n
 Donnellan, Keith, 223n
 doxastic ascent arguments, 143–4
 dreams, 11–19, 26
 Dretske, Fred, 100n, 152n, 266–7, 272n, 290, 332n, 387, 388, 391, 392, 394, 398n, 400n, 445, 455n, 480, 481, 512–3, 515n, 516n, 528n;

Index

- Dretske, Fred (*contd.*)
 perception, 266; relevant alternatives, 80
- Ducasse, C. J., 112, 118n
- Dummett, Michael, 157, 222n, 558, 575
- egocentric rationality, 168, 183, 189–90
- Elgin, Catherine, 515n
- Ellis, R., 554
- Emerson, Ralph W., 460
- empirical knowledge, 120–4, 132, 264; case for, 117; continuous with epistemology, 294, 299–300; *see also* psychology
- empiricism, 120–4; perception, 128–9, 220–1
- ends: rationality of, 196, 568–9, 579–81
- epistemic principles, 249–53
- epistemology: analysis in, 109, 218; armchair, 197, 407–8; as cause of skepticism, 503, 510–1, 513; evolutionary, 314–6; naturalized, 289–90, 303–5, 322–3, 387–8, 410; against naturalized, 305–11, 320–3, 396–8; projects in, 109, 438, 533–4, 556; role of skepticism in, 39–41, 182
- epistemic responsibility, 433–6, 470–1
- Erwin, Edward, 52n
- Erson, Stephen, 467n
- evidence: and intuition, 128; and knowledge, 74–7, 117–8, 149, 161, 226–32, 248, 289, 305–7; undermining, 56, 77, *see also* defeaters
- evidentialism, 167–8, 170
- exclusion, principle of, 12–14, 16–19, 29–30, 80, 275, 282, 389, 482–3, 503
- experience: and coherentism, 227; and knowledge, 108, 112, 226–30; nature of, 228–30, 413–9; perceptual, 30–1, 137, 146, 156–7, 234, 275–6, 293, 404–5, 506; priority of, 536–9; veil of, 19–23, 88, 549–52; veridical, 405
- external world: Cartesian Tradition, 6–23, 325–32; and foundationalism, 26; knowledge of, 6–23, 24–6, 274–6, 293, 338, 413–9, 533, 536–9, 541–2, 544, 551–2; perception 128–9, 220–1, 413–9
- externalism, 200–1, 239–40, 266–7, 270, 278, 283, 387–9, 424–5, 541; belief, 196–9; and coherentism, 395–6; criticism of, 392–8; and internalism, 193, 196–7, 401; and justification, 388, 390–9, 405–12, 424–5, 430; norm, 199–202; in perception, 387–90; reliabilism as a form of, 197, 200–1, 392, 415; and skepticism, 388–90, 401–12; and testimony, 391
- fallibilism, 331, 503–04, 580; and externalism, 504–5, 510–11
- Feigl, Herbert, 118n, 133n, 515n, 569n
- Feinberg, J., 180n
- Feldman, Fred, 243–5, 257n
- Feldman, Richard, 167, 180n, 337, 354, 356–8, 360, 364, 367–8, 370, 370n, 371n, 379–80, 384n, 386n, 445, 455n, 569n, 581n
- felicitous coincidence principle, 55, 61–5
- Ferguson, Kenneth S., 515n
- Firth, Roderick, 98n, 118n, 179n, 222n, 257n, 272n, 311n, 312n, 315, 323n, 353n
- Fodor, J., 223n
- Fogelin, Robert, 241
- Foley, Richard, 152n, 167–8, 169, 190n, 290; egocentric rationality, 168, 183
- Forbes, Graeme, 499n
- Forster, E. M., 40
- Foster, Ian, 456n
- Foster, L., 352n
- Foucault, Michel, 317
- foundationalism, 105, 130–1, 132, 134–6, 168–9, 219, 226–7, 239–40, 249–52, 254–5, 261–5, 290, 538, 541, 548; basic beliefs, 135, 263–6, 302; and coherentism, 105, 143–6; and contextualism, 546–9; criticism of, 120–32; Descartes, 134, 276–7, 302–3; external world problem, 328–32; formal, 106, 141–2, 546–7; perception, 401–5; reconsideration of, 143–6; response to epistemic regress problem, 105, 261–2; semantic, 534, 561–9; and skepticism, 401; substantive, 546–7
- foundherentism, 168–9, 226, 228–35
- Frankfurt, Harry, 243, 256n, 258n
- Fraser, A. C., 456n
- Frege, G., 293, 454, 456n
- Freud, Sigmund, 135, 451, 456n
- Fricker, Elizabeth, 422n
- Fumerton, Richard, 4, 338, 412n
- Gale, Richard, 179n
- Galis, L., 57
- Gettier examples, 55, 68, 73–4, 125, 150, 183–4, 233, 272n, 278; fake barns, 508; Grabit, 55, 61–2, 74, 229; Nogot & Havit, 68–70, 74, 508
- Gettier, Edmund, 55, 65n, 78n, 79–80, 125, 183, 262, 271n, 326, 327, 348, 442, 468, 469, 480, 481, 508, 509, 515n, 517, 519–28, 528n, 529n, 545, 555n
- Gewirth, Alan, 243–5, 256n, 257n
- Gibson, R., 235n
- Ginet, Carl, 97n
- Given, 105, 107–13, 120–4, 130–1, 240, 267–71
- Gödel, Kurt, 292, 321
- Goldman, Alvin, 57, 70–1, 78n, 100n, 101n, 152n, 171, 174, 175, 179n, 180n, 200–1, 222n, 259n, 272n, 278, 285n, 290, 309, 311n, 312n, 316, 322, 337, 352n, 353n, 354, 357–8, 360–2, 365, 367, 368, 370n, 371n, 374–5, 376, 378, 384n, 385n, 387, 388, 390–2, 394, 395, 397, 398n, 400n, 408, 435, 436, 438, 443n, 445, 452, 490, 515n, 555n, 569n, 578, 581n, 582n
- Goodman, Nelson, 85, 118n, 119n, 207, 223n, 235n, 259n, 272n, 412n, 441, 552, 555n, 572–7, 578, 581n
- Grandy, Richard, 499n
- Greco, John, 435, 436, 474n
- Grice, H. P., 542, 555n
- grounding conditions: beliefs, 36–41, 151
- grue, 203, 231
- Guttenplan, Ssmuel, 162n
- Haack, Susan, 106, 168–9, 290, 338; foundherentism, 168–9, 226, 228–35
- habits, 278–9, 361, 362, 369–70, 375–6, 463–4
- Hacker, P. M. S., 569n, 570n
- Hadot, Pierre, 4
- Haldane, E. S., 23n, 191n, 256n, 554n
- Hallen, B., 582n
- hallucination, 404–5
- Hambourger, Robert, 481
- Hankinson, R. J., 4
- Hanson, N. R., 299, 300n
- Hare, R. M., 310, 312n
- Harman, Gilbert, 52n, 55–6, 81, 97n, 142, 152n, 272n, 290, 397, 400n, 515n

- Hartshorne, C., 236n, 352n
 Hauser, Nathan, 467n
 Heath, D., 554n
 Hegel, G. W. F., 124, 131, 314, 317, 348
 Heidegger, Martin, 317
 Heil, John, 171, 174, 179n, 222n
 Heller, Mark, 382–3, 384n, 386n
 Hempel, Carl, 98n, 112, 118n, 231, 235n, 273n, 285n, 311n, 515n
 Hetherington, Stephen, 515n, 516n
 Hintikka, Jaako, 48, 52n
 Hintzman, D., 439
 Hobbes, Thomas, 352n, 459, 466n
 Holism, 157, 198, 290, 303
 Hookway, C., 290, 467n
 Horstmann, R. P., 41n
 Houde, Roland, 119n
 Hubbard, M., 223n
 Hume, David, 33–4, 36–9, 91, 92, 134, 135, 136, 148, 156, 183, 191n, 251, 259n, 293–4, 301, 322, 325, 332n, 388, 398, 407, 418–9, 423n, 454, 530–1, 550, 554n
 Husserl, Edmund, 109, 352n
 hypotheses: skeptical, 11–19, 27–8, 29–32, 49, 51, 56, 86–8, 182, 184, 186, 221, 234, 275, 389–90, 482–3
- immediacy, 27–8, 265, 268, 401–5
 indeterminacy of translation, 161, 295–7, 299
 induction, problem of, 406, 409, 418–9
 inductive beliefs: formation, 37
 inductive inference, 67, 74, 76, 200, 203, 231–2, 362, 368, 406, 409–10
 infallibilism: intuition, 113–4, 116–8, 343–4
 inference to best explanation, 70–2, 74, 231
 inferences: Bayesian, 155; conclusions of, 405–7; inductive, 67, 74, 76, 200, 203, 231–2, 362, 368, 406, 409–10; to the best explanation, 70–2, 74, 231
 infinitism, 105, 139–41, 263
 intellectualist model, 194–5, 217
 internalism, 201–2, 215–9, 239; Cartesian, 182–3; and externalism, 193, 196–7, 201, 337–9; and justification, 192, 337; norm 472–4; and virtue epistemology, 436, 471–4
 internalist approaches: perception, 219–21, 329–32
 intuitions: and evidence, 265, 268–70; self-evident, 125, 131, 249
- invariantism: nonskeptical, 493–5; skeptical, 495–6; warranted assertability objection, 496
 Irwin, T., 285n
- James, William, 174, 179n, 272n, 450
 Johnson, Alexander Bryon, 293, 300n
 Johnson-Laird, P. M., 223n
 justification, epistemic: of cognitive processes, 535, 572–81; conditions, 107–08, 167, 170, 344–52; and contextualism, 549–54; deontologism, 167; description, 107–18; and externalism, 393–8; evaluation of standards of, 558–60, 565–6; evidential, 61, 105, 107–8, 143–9, 170, 192, 262–9; inferential, 126–7, 279–80; and internalism, 192; notions of, 301–2, 340–1, 562–4, 566–9; perceptual belief, 113–7, 220, 329; personal, 228; *prima facie*, 168, 194, 220, 255; procedure, 135; theories of, 134–6, 341–4, 546–9, 556; and truth, 167–9, 344–52, 551; and virtue, 439–40; *see also* regress
- justified true belief (JTB), 55, 60–2, 125, 262; Gettier objections, 55, 68, 73–4, 125, 150, 183–4, 233, 272n, 278
- Kahneman, D., 222n, 581n
 Kant, Immanuel, 24–6, 37, 117, 183, 254, 274, 314, 566
 Kaplan, Mark, 332n, 353n, 481
 Kare, M. R., 300n
 Kenny, Anthony, 191n, 246, 256n, 257n, 258n, 555n
 Kim, Jaegwon, 179n, 289, 312n, 353n
 Kitcher, Philip, 290, 308, 309, 312n
 Klein, Peter, 4, 55, 332n
 Kloesel, Christian, 467n
 knowing-how, *and* knowing-that, 123, 195–6
 knowledge: *a priori*, 117, 131, 240, 254, 262, 265; as act of intellectual virtue, 148, 459–66, 468, 473–4; animal, 239, 280, 282; concept of, 43–4, 58–9, 79, 154, 424–31; defining propositional, 55, 58–9, 62–3, 79, 340–52, 391–2, 445–55, 468–9, 473–4, 503–14; and evidence, 74–7, 117–8, 149, 161, 226–32, 248, 289, 305–07; foundations of, 107–08, 120–4, 135–7, 283, 292, 302; and Gettier objections, 58–9, 79; immediate, 27–8; naturalistic theory of, 33–41, 131, 215–9, 239, 251–2, 255, 266, 290, 331–2, 387–8, 390; noninferential, 120–4, 127–9, 219–21, 263, 269, 278, 290, 330, 401–7, 413–9; object and components of propositional, 58–9, 125; perception and memory, 117; possibility of, 6–23, 25–6; procedural, 194–6; reflective, *see* animal; as reliably true belief, 58–9, 340–52; *see also* empirical knowledge
- Kornblith, Hilary, 171–2, 173–5, 179n, 180n, 222n, 290, 311n, 312n, 462, 466n, 467n
 Kripke, Saul, 52n, 78n, 163n, 443, 499n, 513
 Kruger, L., 41
 Kuhn, T., 299
 Kvanvig, Jonathan L., 400n, 436
 Kvat, Yigal, 96n
 Kyburg, 67, 78n, 98n, 515n
- Lakoff, George, 169n
 language of thought, 206, 209–14
 language: and thought, 115–6
 Leeds, Stephen, 545, 555n
 Lehrer, Keith, 5, 55, 57, 65n, 66n, 78n, 95n, 98n, 106, 142, 152n, 169, 179n, 198, 222n, 235n, 249–51, 254–5, 259n, 260n, 285n, 337, 338, 352n, 398n, 515n, 554n, 555n
 Lepore, E., 423n
 Lepper, M. R., 223n
 Levi, Isaac, 222n
 Lewis, C. I., 106, 107, 109, 115, 116, 117, 118n, 119n, 207, 223n, 227, 268, 271, 285n, 308, 312n, 569n
 Lewis, David, 95n, 96n, 210, 223n, 312n, 480, 481, 515n, 517–28, 528n, 529n, 530n, 533
 limitations, human, 151, 171–2, 179, 183, 205, 234
 Linsky, L., 41n
 Locke, John, 183, 301, 312n, 348, 450, 456n
 Lockwood, M., 52n
 lottery problem, 67, 448, 480–1, 487–9, 508, 514, 520–1, 523, 525–8
 Lovejoy, Arthur O., 116
 luck: epistemic, 60–1, 184, 416–7
 Luper-Foy, Stephen, 499n, 500n, 582n
 Lycan, William, 169
- Mach, E., 318
 MacIntosh, J. J., 41n
 MacIntyre, A., 457, 466n

Index

- Malcolm, Norman, 44–5, 48–50, 52n, 119n
Margalit, Avishai, 83
Marr, David, 379
Marx, Groucho, 81
Marx, Karl, 451, 456n
material world *see* external world
mathematics, examples from, 344, 448, 454, 472; and externalism, 344, 391
Matilal, B., 422n
Maxwell, Grover, 515n
McDowell, John, 169, 338, 420n, 421n, 422n, 423n, 424–31, 431n
McGinn, Colin, 285n, 421n, 555n
McGinn, Marie, 555n
McLaughlin, Brian P., 423n
McMullin, Ernan, 191n
Means, B., 581n
Medin, M., 439, 581n
Meiland, J., 222n
memorial beliefs, 117, 221, 229, 242–3, 247; knowledge, 117, 242–3, 279
Mervis, Carolyn, 169n
metaphysics, 194
Mill, John Stuart, 301, 317
Miller, Caleb, 456n
mind/body dichotomy, 544
modus ponens, 3, 174, 175, 204, 452
modus tollens, 3, 513
Montmarquet, James, 436, 459, 462, 466n
Moore, G. E., 3, 12, 17, 34–5, 47–8, 98n, 240, 274–8, 281, 284, 312n, 314, 482, 494, 495, 500n, 501n, 552
Morgenbesser, S., 300n
Moser, Paul, 57, 169, 523–4, 525, 555n, 569n, 570n
motivation, as basis of intellectual virtue, 458–66
Muirhead, J. H., 41n
Mullally, J. P., 119n
Munitz, Milton K., 515n
Murdoch, D., 284n
Murphy, A. E., 119n
Naess, Arne, 98n
Nagel, Thomas, 529, 530, 554n
Nakhnikian, George, 257n
native belief-forming processes: reliabilism, 175, 194, 446–53
natural and theoretical kinds, 376–9, 406, 443, 542–6, 547
naturalism, 33–41, 131, 215–9, 239, 251–2, 255, 266, 290, 331–2, 387–8, 390; empirical epistemology as, 387–8, 390–1; and justification, 131–2, 301, 331–2
necessary/contingent distinction, 121–2, 218–9
Neurath, O., 136, 273n, 297, 298, 300n
neurological factors: knowledge, 149–50, 184, 186, 206–14, 233
Newell, Alan, 222n
Newman, Lex, 285n
Newton-Smith, William, 191n
Nidditch, P. H., 554n
Nisbett, Richard, 222n, 291, 574, 575, 581n
normativity, 192, 193–7, 289, 301
Nozick, Robert, 5, 56, 78n, 190n, 278, 282, 285n, 388, 392–4, 398n, 400n, 406, 408, 479, 480, 483–6, 489, 491, 494, 499n, 500n
Oakley, I. T., 152n
observation sentences, 298–9
Odegard, Douglas, 332n
O'Hair, Gregory, 98n
OSCAR project, 206–14
Pappas, George, 98n, 152n, 169, 398n
particularism, and methodism, 134, 148
past: beliefs about, 25–6, 403–4
Pastin, Mark, 352
Paxson, Thomas Jr., 55, 65n, 66n, 78n, 515n
Peacocke, Christopher, 420n
Peirce, Charles Sanders, 99n, 234, 236n, 263, 272n, 295–6, 466, 467n, 555n
Penelhum, T., 41n
perception, 105, 117, 120–4, 128–30, 143, 168, 177–8, 209, 229, 276–7, 280, 391; causal theory of, 70–1, 387, 391, 405; doxastic conditions, 117, 368, 430; nature of, 361–2, 379, 388, 390–1, 409, 413–6, 418–9; principles of, 220, 329–30; testimony and, 178, 391; as unreliable source, 394
perceptual and memory knowledge, 105, 117
performative, 125–6; performative fallacy, 126
Perry, John, 210, 223n
person: rationality of, 305–7; theory or, 419
Pettit, Philip, 423n
phenomenalism, 107–8, 113, 118, 207, 303–4
phenomenological reduction, 294–6, 302–5
philosophy: *a priori* investigation, 197, 218, 320–3
physical objects: perception, 128–9, 137, 220–1, 228–30, 413–9
Piaget, Jean, 447, 455n
Plantinga, Alvin, 57, 106, 169, 191n, 291, 356, 364, 365, 370n, 371n, 435–6, 437, 440–1, 455n
Plato, 59, 60, 271n, 566; Knowledge, 60; *Meno*, 59, 60, 93; *Theaetetus*, 59, 271n, 567
Polanyi, M., 299
Pollard, C. J., 223n
Pollock, John L., 57, 65n, 96n, 168, 169, 180n, 260n, 312n, 360, 364, 370n, 371n, 384n, 442, 445, 447, 455n, 474n, 569n; OSCAR, 206–14
Popkin, Richard, 5, 98n, 241
Popper, Karl, 348, 555n
possibility *and* possibilities, 27–8, 31–2, 41, 154, 245–6, 411, 480, 505–14
Post, John F., 152n, 569n
pragmatism, 115
Price, H. H., 23n, 122
Prichard, H. A., 245–7, 250, 257n
principles, epistemic: identification of, 132, 192; and psychology, 196–7, 294–5
priority, epistemic, 297–8, 536–8, 540–1, 547–9
probabilism: epistemic justification, 67–8, 110, 127, 198, 252
probabilistic arguments, 67–8, 72–3, 127, 198
process *and* processes, 176–8, 194; cognitive, 170, 175, 347; and generality problem, 315–6, 337–8, 354–6, 357–70, 372–84, 406–7, 409–10; reliable, 345–7, 363–7, 369–70, 372–84, 406
proof, conditions for rigorous, 24, 274
proper function, 435–6, 446–55; and design plan, 435–6, 451–3, 455
propositions, disqualifying, *see* defeaters
propositions, true: in knowledge, 167–9, 183–5, 279–80
psychological factors: knowledge, 149–50, 184, 186, 233
psychology: and analysis of epistemic concepts, 218, 402–3, 407–8; and epistemic principles, 196–7, 360–1; relation to epistemology, 289, 403, 419; role in epistemology, 196–7, 234, 289, 297–8, 306, 308–10, 365–7, 369–70, 377–80

- Putnam, Hilary, 155, 162n, 163n, 191n, 222n, 223n, 289–90, 498n, 525, 580
- Pyrrho of Elis, 92, 182
- Pyrrhonian problematic, 105, 239–40
- Pyrrhonian skepticism, 92, 182
- Quine, W. V., 35, 36, 41n, 119n, 142, 157–61, 162n, 235n, 259n, 284n, 289–90, 291, 300n, 303–10, 311n, 312n, 320–2, 324n, 331, 348, 352n, 388, 398, 410, 411, 412n, 527–8, 542–5, 547, 555n, 564; analytic/synthetic distinction, 157, 298, 542–4; naturalism, 35, 36, 289–90, 303, 305–08, 320–3
- Quinton, Anthony, 261, 262, 268–9, 271, 271n, 273n
- Radford, Colin, 515n
- radical interpretation, 159–61
- radical translation, 159–61
- Ramsey, F. P., 448
- Rapaport, W. J., 223n
- rationality, 28, 167, 183; egocentric, 168, 183; human, 151, 171–2, 179, 183, 205, 234; personal, 305–7
- Rawls, John, 572, 575, 581n
- realism: epistemological, 536–54; internal, 155; metaphysical, 155, 290
- realism, direct, 199, 219–20; perception, 199
- reasons and reasoning: epistemologically circular, 94–5, 185, 187–8, 232, 281–4, 290, 410; limitations of human, 151, 171–2, 179, 183, 205, 234; in OSCAR, 209–15; from percepts, 107–8, 113–4, 117–8, 207, 219–21, 414–5, 418–9; and rationality, 126–7, 203–4, 213–5
- reduction, 292–5, 303; phenomenological, 294–6, 302–5
- reflective equilibrium, 572–7
- reflexivity, 94–5, 185, 187–8, 232, 281–4, 290, 410
- regress, 78, 129, 194, 232, 264
- regress argument, epistemic, 261–2; coherentism, 144; foundationalism, 261; *see also* Agrippa's Trilemma
- Reichenbach, Hans, 110, 114–5, 118n
- Reid, Thomas, 235n, 498
- relativism, 202, 204, 234, 281, 299, 316–9, 534, 579–80; conceptual, 568–9; and contextualism, 549–54; cultural, 316–20
- relativity, cultural, 182, 579–80
- relevant alternatives (RA), 389, 485, 487–9, 508–14, 518, 520–8
- reliable sources, 8–10, 283, 345–7, 363–7, 369–70, 372–84, 406; induction as, 408, 441; memory as, 408, 441, 509; perception as, 123–4, 408, 430, 441, 509; testimony as, 509
- reliabilism, 167, 175–9, 183, 197, 200–2, 278; analyses of knowledge, 345–52, 363–7, 369–70, 372–84, 387, 392; criticism of, 354–6; defense of, 357–70; and internalism, 415; and justification, 404, 409, 427, 559, 578; native processes, 175, 194, 446–53; and proper function, 453–5
- reliabilist approach: perception, 390, 406; testimony, 350, 391
- Rescher, Nicholas, 95n, 106, 142, 272n
- responsibility, epistemic, 172, 192–3
- Rey, G., 581n
- Riding, A., 443
- Roberts, Robert, 457, 466n
- Rorty, A. O., 464, 467n
- Rorty, Richard, 156, 162n, 191n, 282, 285n, 311n, 312n, 317, 323, 423n, 525
- Rosch, Eleanor, 569n
- Ross, Glenn, 5
- Ross, L., 222n, 223n
- Ross, W. D., 23, 191n, 256n, 554n
- Roth, M. D., 5, 57
- Routley, R., 581n
- rule-following, 122, 168
- Russell, Bertrand, 27–8, 29, 50, 52n, 60, 109, 260n, 272n, 275–6, 284n, 293–4, 308, 312n, 314, 456n, 515n, 566
- Ryle, Gilbert, 352n
- Salmon, W., 582n
- Schaffer, M. M., 439
- Scheffler, Israel, 119n, 272n
- Schiffer, Stephen, 223n, 481
- Schilpp, Paul Arthur, 119n
- Schlick, M., 273n
- Schmitt, Frederick, 180n, 311n, 354, 370n, 381–2, 384n, 386n, 569n
- science: acceptance of probability of error, 187; continuous with epistemology, 289–90; philosophy of, 188; rationality of scientists, 208; reliability of, 187; role in epistemology, 289–90, 292, 338, 376–80, 438–9
- scientia*: *see cognitio*, Descartes
- Scribner, S., 581n
- Scriven, Michael, 118n, 133n
- Selby-Bigge, L. A., 554n
- self-evidence: immediate, 255; and justification, 127, 341–2
- self-presenting propositions, 343–4
- Sellars, Wilfrid, 105, 106, 118n, 119n, 142, 152n, 249–52, 259n, 273n, 284n, 285n, 338, 403, 412n, 413–8, 420n, 423n, 424–6, 429–30
- semantic ascent, 322
- Senor, Tom, 443n, 456n
- sense-impressions, 107, 118
- sensory and perceptual experience, 30–1, 137, 146, 156–7, 234, 275–6, 293, 404–5, 506; conceptual formulations, 219–21, 228–30, 361–2, 379, 388, 409, 413–9; veil of, 19–23, 88, 549–52
- Sextus Empiricus, 98n, 281
- Shope, Robert, 57, 101n, 582n
- Siegel, Harvey, 525
- Simon, Herbert, 222n
- Singer, Marcus, 235n
- skeptical arguments: evaluating, 86–8, 182–90, 401–5, 479–80, 482–6, 487–94, 496, 498
- skeptical hypotheses, 11–19, 27–8, 29–32, 49, 51, 56, 86–8, 182, 184, 186, 221, 234, 275, 389–90, 479, 482–3; brain in a vat hypothesis, 56, 221, 359, 387–8, 469, 474, 480–94; Evil Deceiver, 234, 245–6, 326, 359, 387–8, 469, 474, 480–94; fake barns, 80, 490
- skepticism, 3–4, 42–3, 48–52, 56, 86, 131, 156–8, 161–2, 189, 221, 234–5, 263, 270, 274, 281; Academic, 114; Agrippa's Trilemma, 281; Ancient, 114, 281; Cartesian Arguments, 11–19, 533; and contextualism, 492–5, 509–14, 518–9, 522–3, 526–8, 549–54, 579–80; and foundationalism, 538–41; and knowledge, 65; and naturalism, 33–41; Pyrrhonian, 92, 182; role in epistemology, 39–41, 182, 185–7, 189
- Skyrms, Brian, 78n, 582n
- Slote, Michael, 98n
- Slovic, P., 222n, 581n
- Smith, David Woodruff, 223n
- Smith, E. E., 439, 440, 581n
- Smith, Holly, 443
- Sobel, J. H., 96n
- Sober, E., 311n

Index

- social factors: knowledge, 150, 338–9, 425–6, 428–30
- social standards: knowledge related to, 150, 441–2, 568–9, 579–80
- Socrates, 60, 93, 109
- Sodipo, J. O., 582n
- solipsism, 317–9
- Sosa, Ernest, 5, 57, 60, 65n, 66n, 106, 169, 191n, 224n, 235n, 239–40, 241, 272n, 291, 312n, 332n, 338, 352n, 384n, 386n, 436, 437, 439, 442, 443n, 466n, 468–73, 474n, 475n, 481, 525, 555n, 569n
- space of reasons, 413–9; as socially articulated, 424–43; *see also* Sellars
- speaker-sensitive rules, 520–1, 522–8
- Spedding, J., 554n
- Spinoza, Baruch, 459
- Stalnaker, Robert, 95n, 96n, 515n, 516n
- standards, epistemic, 569, 579–80; contextualism, 549–54
- states: internal, 193; non-doxastic, 220–1
- Stein, Edward, 291
- Stich, Stephen, 291, 352n, 524–5, 581n
- Stine, Gail, 80, 97n, 488, 515n
- Stout, A. K., 256n
- Stoothoff, R., 284n
- Strachey, James, 456n
- Strawson, P. F., 4, 5, 241, 542, 555n, 582n
- Stroud, Barry, 3, 5, 34–6, 240, 241, 283, 284n, 285n, 495, 500n, 501n, 515n, 526, 528, 529, 554n
- subject, knowing: nature of, 278–9
- subjunctive conditionals, 79–82, 387
- supervenience, 85, 142–3, 253, 289–90, 310–1, 332
- Swain, Marshall, 98n, 152n, 278, 285n, 354, 370n, 397, 400n
- Swanson, J. W., 352n
- Swartz, Robert J., 273n
- Swinburne, Richard, 456n
- Tarski, Alfred, 155, 159, 321, 322
- Teichman, J., 515n
- testimony, 229, 281, 391, 393, 407, 509
- Thagard, Paul, 171–2, 179n
- theories, epistemological, 60, 215–9, 281–4, 408–11
- thermometer view, 122, 317, 392–4, 402, 415, 425
- Tiles, Jim, 467n
- Tiles, Mary, 467n
- Tomberlin, James, 235n, 400n
- tracking *and* sensitivity, 56, 82, 278, 282, 387, 392, 402, 406, 408, 469–80, 484–95, 496
- transcendental arguments, 36, 40–1
- Trianosky, Gregory, 457, 466n
- trust, 188–9, 398–9; role in knowledge, 190, 399
- truth, 154, 167–9, 172, 183–5, 234, 246–8, 265, 269, 279–80, 544–5; coherence theories of, 110–1, 154; conditions, 154; correspondence, 154–5, 314–5, 322; deflationism, 544–5; and justification, 167–9, 172, 183–5, 279–80, 282; objective, 544–5; as rational acceptability, 315; and realism, 544–5
- truth condition analysis, 159–61, 202–4
- truth conditions, 202, 213–4
- Tversky, A., 222n, 581n
- Uehling, Theodore E. Jr., 456n
- Ullian, J., 235n, 259n, 352n
- underdetermination arguments: and skepticism, 295–7
- Unger, Peter, 4, 5, 52n, 98n, 272n, 278, 285n, 481, 495, 498n, 499n, 500n, 501n, 515n; absolute terms, 46–7; dogmatism, 47–8, 50–2; invariantism, 495–6
- Urmson, J., 132n
- Van Cleve, James, 152n, 191n, 222n, 239–40, 312n
- van Fraassen, Bas, 191n
- verbal behaviorist model, 129–30, 132
- virtue: definition, 436, 457–8, 468; and coherence, 279–81; epistemology, 278–81, 435, 6, 468–74; ethics, 148, 457–8; intellectual, 148, 278–9, 458–66, 468–9
- visual object recognition: reliability of, 129–32, 219–21, 329
- voluntarism, doxastic, 178
- Waismann, F., 569n
- Wallis, Charles, 375, 384n, 385n
- warrant, 65, 278–9, 283, 435–6, 445–6, 447–50, 453–5
- warranted assertability maneuvers (WAM), 495–6
- warranted assertability objection, 496
- Wason, P. C., 223n
- Weiss, Paul, 118n, 236n
- Werkmeister, W. H., 119n
- Wettstein, Howard, 456n
- Whitehead, A. N., 352n
- Wild, John, 118n
- Williams, Bernard, 191n, 421n
- Williams, Michael, 5, 152n, 332n, 515n, 523, 525, 555n
- Wittgenstein, Ludwig, 34, 36, 38–40, 41n, 98n, 108, 109, 142, 245–6, 257n, 273n, 297, 316, 319, 423n, 481, 531, 545, 546, 548, 549, 550, 555n, 563, 569n, 579
- Wood, Ledger, 117, 119n
- Woozley, A., 515n
- Wright, Crispin, 332n, 333n, 420n, 422n, 481, 553–4
- Yilmaz, Hüseyin, 300, 300n
- Zagzebski, Linda, 57, 436, 437
- Zimmerman, Dean, 456n